



HAL
open science

Amélioration des chaînes de traitement d'images satellites

Juliette Miranville

► **To cite this version:**

Juliette Miranville. Amélioration des chaînes de traitement d'images satellites. Informatique [cs]. 2024. hal-04584953

HAL Id: hal-04584953

<https://hal.univ-reunion.fr/hal-04584953v1>

Submitted on 23 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

UNIVERSITÉ DE TECHNOLOGIE DE COMPIÈGNE

GÉNIE INFORMATIQUE

Rapport de stage TN09

Amélioration des chaînes de traitement d'images satellites

Lieu de stage : Station SEAS-OI
40 avenue de Soweto, Saint-Pierre, 97410, La Réunion

Réalisé par :
Mme. MIRANVILLE Juliette

Encadré par :
M. MOUQUET Pascal (IRD)
M. BEN BELGACEM Faker (UTC)

Financé par : l'IRD et DEAL Réunion dans le cadre du projet BioEOS.



Août 2023 - Février 2024

Remerciements

Je tiens à remercier toutes les personnes qui ont joué un rôle essentiel dans la réussite de mon stage. Cette expérience a été vraiment enrichissante, et je suis reconnaissante envers chacun de vous pour votre soutien, vos conseils et votre encouragement.

Un grand merci à Pascal Mouquet, mon tuteur de stage, pour son expertise, sa patience et sa disponibilité. Ses conseils avisés et ses connaissances ont vraiment contribué à mon apprentissage et à mon développement professionnel.

Je souhaite également exprimer ma gratitude envers toute l'équipe de SEAS-OI. Votre accueil chaleureux, votre collaboration et l'ambiance conviviale ont rendu mon expérience de stage exceptionnelle. Travailler avec des collègues aussi inspirants a été un réel plaisir.

Merci à mes professeurs à l'Université de Technologie de Compiègne pour leur encouragement et leur soutien tout au long de ma formation. Un merci spécial à Maryam Ly, Ben Belgacem Faker et à tous les autres qui ont contribué de près ou de loin à la réussite de mon stage.

Cette expérience a été vraiment formatrice, et je tire de précieuses leçons et souvenirs qui seront utiles tout au long de ma carrière.

Merci à tous pour votre générosité et votre confiance.

Table des matières

Introduction générale	1
1 Présentation de l'institut d'accueil	3
1.1 Historique	4
1.1.1 L'IRD	4
1.1.2 La station SEAS-OI	4
1.1.3 L'UMR Espace-Dev	6
1.2 Présentation générale	6
1.3 Présentation de l'équipe d'accueil	7
1.4 Méthodologie	7
1.5 Outils utilisés	8
1.5.1 GitLab	8
1.5.2 Python	8
1.5.3 X2Go	8
1.5.4 Utilisation de différents serveurs	9
2 Ma mission	10
2.1 L'outil Sen2Chain	11
2.1.1 Données de Sentinel-2	11
2.1.2 La chaîne de traitement Sen2Chain	14
2.2 Objectifs de ma mission	17
2.2.1 Mise en place du cahier des charges	18
2.3 Planning	18
3 Déroulement du projet	20
3.1 Appropriation du sujet	21
3.2 Mise à jour de l'outil d'extraction des séries temporelles	21
3.2.1 La classe TimeSerie	21
3.2.2 Objectifs	22
3.2.3 Arborescence d'extraction	23
3.2.4 Mise en place d'une somme de contrôle	24
3.2.5 Données brutes	25
3.2.6 Données compilées	27
3.2.7 Présentation du travail effectué	32
3.2.8 Intégration du lissage temporel	33
3.2.9 Fichier log	34
3.2.10 Modification de la classe Produit	34
3.2.11 Mise en production	35
3.3 Mise en place d'une routine d'extraction	35
3.3.1 La classe Job	35
3.3.2 Objectifs	36

3.3.3	Classe JobExtraction	37
3.3.4	Fichier de configuration	37
3.3.5	Utilisation du cron	37
3.3.6	Copie vers un serveur distant	38
3.4	Mise à jour des produits	38
3.4.1	Classe Tile	38
3.4.2	Objectifs du travail demandé	38
3.4.3	Gestion des doublons	39
3.4.4	Classe download EODAG	40
3.4.5	Gestion des téléchargements	40
3.5	Developpement du rover centipede	41
3.5.1	Présentation du réseau Centipede RTK	41
3.5.2	Objectif	42
3.5.3	Déroulement	42
3.6	Participations à d'autres activités liées à la télédétection et à l'informatique	43
3.6.1	Rencontres Géomatiques de la Réunion (RGR - 2023)	43
3.6.2	Restitutions du Projet G2Oi	43
3.6.3	Assemblée générale de l'IRD	44
	Bilan final	45
3.7	Perspectives	46
3.8	Bilan personnel	46
	Annexe	48
	Bibliographie	50

Résumé

Les stages en entreprise ou laboratoire font partie intégrante de la formation d'ingénieur de l'Université de Technologie de Compiègne (UTC).

Dans ce cadre, j'ai réalisé un stage de 24 semaines à la station de Surveillance de l'Environnement Assistée par Satellite pour l'Océan Indien (SEAS-OI). L'objectif principal était l'exploration de nouvelles fonctionnalités et le développement informatique en vue d'améliorer les chaînes de traitements de données spatiales existantes.

Ce rapport est consacré au compte rendu de ce travail; plus particulièrement aux missions qui m'ont été confiées, aux difficultés rencontrées et les solutions proposées ainsi qu'aux apprentissages et bilans acquis.

Introduction générale

Cadre du stage

Le troisième semestre du programme de formation en ingénierie à l'Université de Technologie de Compiègne (UTC) est dédié à un stage en tant qu'assistant ingénieur. Cette immersion professionnelle, qu'elle se déroule dans le secteur privé ou public, revêt une importance cruciale dans le parcours académique. L'objectif premier de cette expérience est d'amener les étudiants à affronter les réalités du monde du travail, à comprendre ses contraintes et enjeux, tout en mettant en œuvre les connaissances acquises jusqu'à présent. Ce stage vise ainsi à offrir des enseignements concrets tirés de l'application des compétences académiques.

Mon stage s'est déroulé à la station SEAS-OI, une plateforme d'acquisition et de traitement d'images haute résolution portée et financée par la Région Réunion, l'Etat, l'Institut de Recherche pour le Développement (IRD), l'Université de La Réunion, et soutenu par l'Europe. Des activités de recherche sont animées par l'Unité Mixte de Recherche (UMR) Espace-Dev à travers de nombreuses collaborations avec des institutions et des acteurs de La Réunion et de l'océan Indien.

Contexte du stage

Les images de télédétection offrent une précision spatiale et temporelle exceptionnelle, contribuant ainsi à une meilleure compréhension de notre environnement et du fonctionnement des écosystèmes. Au fil des années, l'UMR Espace-Dev a consacré ses efforts à l'élaboration de méthodes pour surveiller les dynamiques spatiales. Plus précisément, la branche réunionnaise de cette unité établie à la station SEAS-OI, s'est engagée dans le développement de chaînes de traitement d'images destinées à l'observation de la Terre, en vue de caractériser divers phénomènes.

Des programmes de recherche récents ont abouti à la mise en place de chaînes de traitement opérationnelles d'images satellites permettant la mise en service d'outils de production et de partage d'informations environnementales qualifiées dans des domaines tels que la santé et le climat ou encore les risques naturels et les dynamiques côtières. Ces outils, parmi lesquels certains ont reçu des labels de qualité, reposent sur des langages de programmation, des bibliothèques et des données en accès libre, garantissant ainsi leur pérennité. Il est essentiel de maintenir ces outils à jour, d'améliorer leurs fonctionnalités existantes et d'intégrer de nouvelles capacités pour garantir leur durabilité et répondre aux besoins variés des utilisateurs.

Dans ce contexte, ce stage s'inscrit dans le cadre du projet de recherche BioEOS¹, visant à relever le défi de l'obtention de données fiables et pertinentes pour suivre la biodiversité à grande échelle en utilisant principalement des images de télédétection. Plus précisément, BioEOS a pour objectifs de caractériser la dynamique spatiotemporelle de la biodiversité grâce à l'analyse de

1. BioEOS : <https://www.spaceclimateobservatory.org/fr/bioeos>

séries temporelles d'images satellites multicateurs et de cartographier les changements et de produire des indicateurs pour évaluer l'impact des changements globaux sur la biodiversité des écosystèmes côtiers. Ce stage a bénéficié d'un cofinancement de la DEAL Réunion dans le cadre du volet Réunion du projet BioEOS, et de l'IRD.

Objectifs et déroulement

Pendant la période de stage, l'objectif principal était de contribuer au développement de l'outil Sen2Chain² en collaboration avec les développeurs des chaînes de traitement. Mon rôle consistait à proposer des améliorations, à ajouter de nouvelles fonctionnalités et à intégrer ces modifications de manière opérationnelle dans les outils déjà existants au sein de la station. Outre ces responsabilités, j'ai également eu l'occasion de participer à diverses activités liées à la télédétection et à l'informatique au sein de la station et de l'UMR. Cela incluait notamment l'exploitation opérationnelle du réseau ouvert Centipède³ de positionnement satellite centimétrique GNSS RTK.

Organisation du rapport

Dans une première partie, l'entité d'accueil et l'équipe au sein de laquelle le stage s'est déroulé seront brièvement introduites. L'objectif est de fournir un contexte clair pour la suite du rapport.

Le deuxième chapitre se consacrera à la mission au sein de l'équipe d'accueil. Cette partie permettra de mieux comprendre le travail accompli au cours du stage et son importance pour l'organisation d'accueil.

Le troisième chapitre se concentrera sur les réalisations durant le stage. Cette section détaillera le travail accompli et les résultats concrets apportés aux projets.

2. Le Framagit de l'outil Sen2Chain : <https://framagit.org/espace-dev/sen2chain>

3. Plus d'informations sur Centipede sont disponibles à l'adresse suivante : <https://docs.centipede.fr/>

Chapitre 1

Présentation de l'institut d'accueil

1.1 Historique

1.1.1 L'IRD

L'Institut de Recherche pour le Développement est un organisme de recherche public français pluridisciplinaire reconnu internationalement qui, depuis près de 80 ans, s'engage dans des partenariats équitables avec les pays du Sud et dans les Outre-mer français. L'Institut est un établissement public français placé sous la double tutelle du ministère de l'enseignement supérieur de la recherche et de l'innovation et de celui de l'Europe et des affaires étrangères. Il porte une démarche de recherche, d'expertise, de formation et de partage des savoirs au bénéfice des territoires et pays qui font de la science et de l'innovation un des premiers leviers de leur développement.

L'IRD s'appuie sur son réseau de délégations régionales en France métropolitaine et de représentations en Outre-mer et à l'étranger pour organiser ses collaborations.



FIGURE 1.1 – Délégations régionales et représentations de L'IRD dans le monde

La Réunion est un territoire européen insulaire et tropical, elle fait partie de la région Afrique Australe, Afrique de l'Est et océan Indien, dont les autres représentations sont basées en Afrique du Sud, au Kenya, à Madagascar et aux Seychelles. Cela permet de créer un environnement propice au développement d'activités de recherche, de développement et d'innovation. L'IRD s'est installé à La Réunion en 1987 avec le premier programme de suivi de l'environnement marin lié au développement de la pêche thonière française dans l'océan Indien. Au fil des années et des programmes, les équipes se sont étoffées pour répondre aux différentes problématiques du territoire et de la région. La représentation de la Réunion, vient également en appui des activités de l'IRD à Mayotte et aux Iles Eparses.

1.1.2 La station SEAS-OI

La station SEAS-OI, est une plateforme de réception d'imagerie satellite. Elle est opérationnelle depuis juillet 2012 et est le résultat d'un projet qui a débuté en 2004. Au cours de cette période, les porteurs de projet ont travaillé sur la communication, la planification et la mise en forme du programme. En 2008, une convention quadripartite a été signée entre la Région Réunion, l'État, l'IRD et l'Université de La Réunion, établissant les bases de la coopération. Des séminaires ont été organisés en 2009 pour comprendre les besoins des futurs utilisateurs. La

période 2009-2011 a été consacrée au montage de la plateforme SEAS-OI, avec la signature de contrats industriels et la création de la structure adéquate. La station a été livrée en septembre 2011 et inaugurée en octobre 2012. En 2012, un accord-cadre pour la gouvernance du pôle en télédétection a été signé entre les quatre partenaires, marquant une étape importante dans le projet.

La station est située à Saint-Pierre de la Réunion. La station a la capacité de recevoir dans un rayon de 2500 km autour de La Réunion des images satellites optiques et radar haute résolution. SEAS-OI couvre toutes les îles de l'océan Indien occidental, une grande partie du Mozambique, le Swaziland, une partie de la Tanzanie, du Zimbabwe et de l'Afrique du Sud, ainsi que les deux grandes zones maritimes que sont l'océan indien occidental et le canal de Mozambique.

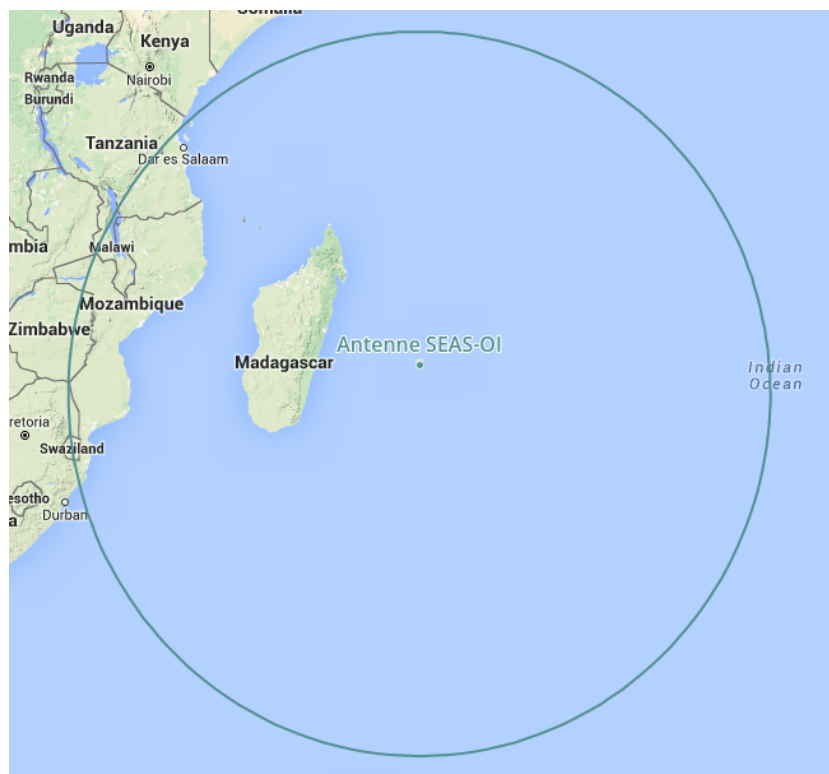


FIGURE 1.2 – Zone d'acquisition de l'antenne [Crédits : SEAS-OI]

Elle a reçu des images SPOT4 de juillet 2012 à la désorbitation du satellite en janvier 2013 et des images SPOT5 de juillet 2012 à mars 2015. La réception directe est arrêtée depuis décembre 2015, mais la capacité de réception est fonctionnelle, le dispositif technique est aujourd'hui maintenu en condition opérationnelle par l'Université de La Réunion. Depuis 2017, la station récupère indirectement les images des satellites du programme européen Copernicus pour alimenter les chaînes de traitement opérationnelles et générer des informations environnementales de haute qualité. Ces informations sont utilisées dans divers domaines, y compris la santé, le climat, les risques naturels et les dynamiques côtières.

Une fois ces images traitées, elles sont envoyées sur un serveur qui permet l'archivage, la sauvegarde et le requêtage de la base de données image. Le catalogue en ligne de SEAS-OI est basée sur cette infrastructure serveur.

1.1.3 L'UMR Espace-Dev

Dans le sud-ouest de l'océan Indien, l'UMR Espace-Dev est implantée au sein de l'Université de La Réunion. L'équipe est hébergée dans les locaux de l'IRD (Institut de Recherche pour le Développement) sur le campus Universitaire du Moufia et à la station SEAS-OI sur le campus de Terre Sainte. Depuis les années 1990, l'UMR et la structure dont elle est issue (l'Unité Espace de l'IRD) mènent dans le sud-ouest de l'océan Indien des travaux à La Réunion, Mayotte, Madagascar, les Comores, les îles Eparses et les espaces maritimes associées (canal du Mozambique, Sud-Ouest de l'océan Indien). Ces travaux sont largement empreints de télédétection depuis la création du réseau SEAS (IRD) en 1991 à La Réunion. Ils s'inscrivent aujourd'hui dans les projets et réseaux nationaux tels que GEOSUD-GEODEV (Théia) et Kalideos (CNES). En parallèle, avec ses partenaires l'UMR anime des observatoires de l'environnement (ex : SNO DYNALIT et usages de littoraux avec l'OSU-Réunion, dynamiques d'occupation du sol à l'échelle des Mascareignes).

1.2 Présentation générale

Au sein de SEAS-OI, l'IRD, l'Université de La Réunion, et la Région Réunion (Conseil Régional de La Réunion) mettent en oeuvre des projets de coopération scientifique et opérationnelle dans le domaine de la télédétection. Cette station a pour objectif d'acquérir, de traiter et de diffuser des images satellites afin de soutenir la compréhension, l'analyse et la gestion intégrée de l'environnement et des territoires dans la région du bassin ouest de l'océan Indien. Les données acquises par la station ont vocation à développer des applications dans les domaines de l'aménagement du territoire (urbanisation, foncier, gestion intégrée des zones côtières, agriculture, gestion forestière), de l'environnement marin et côtier (trafic, pêche, pollution), des risques naturels, de la biodiversité, de la santé et des changements climatiques. En collaboration avec les partenaires tels que l'IRD, l'Université de La Réunion, l'État et la Région, un pôle d'expertise en télédétection a été développé autour de la station. Ce pôle a quatre principaux objectifs :

1. Recherche scientifique : Il vise à réaliser des cartographies dynamiques, à élaborer des indicateurs et à créer des modèles pour une meilleure compréhension de la région.
2. Formation : Le pôle s'engage à proposer une formation initiale et continue en géomatique et en télédétection afin de renforcer les compétences dans ce domaine.
3. Innovation technologique : Il encourage le développement de services opérationnels et de la surveillance environnementale en utilisant les données satellitaires.
4. Coopération régionale : Le programme s'inscrit dans une démarche de coopération régionale dont le but est le codéveloppement des pays et territoires de la zone océan Indien, en relation avec la Commission de l'océan Indien, regroupant 5 pays (Comores, France-Réunion, Madagascar, Maurice et les Seychelles). La position géographique particulière de La Réunion et ses relations historiques avec les autres îles du sud-ouest de l'océan Indien et des pays d'Afrique australe et orientale, favorisent le développement des actions de coopération régionale. Les secteurs d'intervention recouvrent des préoccupations communes telles que le développement durable, l'intégration économique régionale et le développement humain et la solidarité internationale.

Les activités de recherche sont animées par l'Unité Mixte de Recherche (UMR). Elle organise des séminaires dans l'amphithéâtre de SEAS-OI, abordant des thématiques liées à l'observation spatiale articulés autour de quatre axes majeurs. Tout d'abord, l'étude de l'occupation des sols vise à mieux comprendre les usages du territoire pour une planification efficace. En parallèle,

la santé environnementale explore les interactions entre l'environnement et la santé publique. L'UMR se penche également sur l'interface littoral, analysant les dynamiques spécifiques entre les zones terrestres et marines. Enfin, la recherche sur les risques naturels cherche à anticiper et à prévenir des événements tels que les inondations et les mouvements de terrain, contribuant ainsi à la gestion des catastrophes. De plus, l'UMR répond à des appels à projets visant à améliorer la couverture de l'imagerie satellitaire dans le Sud-Ouest de l'océan Indien, incluant des initiatives telles que Kalidéos 2 et la production de produits Sentinel-2. En outre, elle favorise la collaboration en accueillant des chercheurs d'autres institutions, comme le Cirad, le CHU, et diverses universités. Cette synergie entre les acteurs de la recherche en télédétection et information spatialisée à La Réunion contribue à renforcer le réseau de recherche dans la région.

Des activités de valorisation sont portées par la Région Réunion qui s'engage dans divers projets collaboratifs, illustrés par des initiatives telles que GRIMA, Orthobase et PROFETS.¹

1.3 Présentation de l'équipe d'accueil

L'équipe SEAS-OI, sous la direction de Rodolphe Devillers, Directeur de Recherche à l'IRD et affilié à l'UMR Espace-Dev, regroupe des chercheurs et professionnels issus de diverses institutions, collaborant pour des projets liés à l'environnement et la géomatique dans la région océan Indien.

Du côté de l'IRD, des chercheurs, ingénieurs et chargés de recherche participent activement à des projets variés, couvrant des domaines tels que la géomatique, la télédétection, la sociologie, et l'environnement littoral. En parallèle, l'Université contribue avec des experts en informatique, géographie, géomatique, environnement littoral, et sociologie. Les enseignants-chercheurs et ingénieurs d'études de l'Université apportent une expertise complémentaire à l'équipe, participant à la résolution des défis environnementaux dans la région océan Indien. La collaboration avec la Région Réunion, notamment à travers une cellule SIG et des experts en télédétection, renforce les capacités de SEAS-OI,

En hébergement, les membres de la société CLS sont impliqués. Des doctorants et stagiaires enrichissent également l'équipe. CLS, est une société internationale affiliée au CNES et à CNP. Elle excelle dans cinq domaines : gestion des pêches, surveillance environnementale, sécurité maritime, mobilité, et énergies/infrastructures. Spécialisée dans les services satellitaires, CLS contribue à la protection de la planète par la localisation, la collecte de données environnementales, et l'observation des océans et terres.

Ensemble, l'équipe de SEAS-OI aborde de manière collaborative les défis environnementaux de la région océan Indien.

1.4 Méthodologie

Les directives et les orientations du travail étaient principalement définies au travers de discussions internes et de prises de décision au sein de l'équipe. L'absence d'une structure client-formel dans ce contexte a éliminé la nécessité de consultations externes. Les décisions importantes et les ajustements nécessaires étaient plutôt discutés et décidés lors de réunions dédiées avec mon tuteur. De plus, la validation du code intervient par le biais d'une série de tests et de la mise en application du code, contribuant ainsi à consolider la rigueur du processus de développement.

1. Plus d'informations sur ces projets : <http://www.seas-oi.org/web/guest/valorisation>

Cette approche, bien que partageant des similitudes avec la méthodologie AGILE, se distinguait par son processus de révision interne du projet. La flexibilité des délais a également encouragé une autonomie prononcée dans la gestion du projet, facilitant ainsi une adaptation continue aux évolutions du projet.

1.5 Outils utilisés

Pendant la période du stage, divers outils ont été utilisés. Voici une liste, bien que non exhaustive, de certains d'entre eux.

1.5.1 GitLab

Le développement du projet a été facilité par l'utilisation de GitLab, une plateforme collaborative robuste qui offre des fonctionnalités avancées pour la gestion du code source. GitLab a permis une coordination efficace du travail d'équipe, avec des outils intégrés pour la collaboration, le suivi des problèmes, et la gestion des pipelines d'intégration continue.

Au cœur de cette infrastructure, l'outil Git a été central pour le contrôle de version. Git offre une approche décentralisée qui permet à chaque membre de l'équipe de travailler de manière indépendante sur sa branche de code, facilitant ainsi le développement simultané de fonctionnalités sans conflits majeurs. La gestion des branches, la fusion aisée des modifications et la possibilité de revenir à des versions antérieures ont contribué à la flexibilité et à la stabilité du processus de développement.

L'instance Framagit, hébergée par l'association Framasoft, a été choisie en tant que déploiement spécifique de GitLab. Framagit a joué un rôle clé en offrant un environnement libre, décentralisé et sécurisé pour le stockage du code source. Cette plateforme a favorisé une collaboration transparente en fournissant une visibilité claire sur les contributions individuelles, les commentaires et les évolutions du code.

Le choix de Framasoft et de l'instance Framagit reflète également une préférence pour une approche ouverte et inclusive dans le domaine des outils collaboratifs. En optant pour des solutions libres et décentralisées, l'équipe a mis en avant les valeurs fondamentales de l'open source, contribuant ainsi à créer un environnement de travail transparent, sécurisé et propice à la collaboration.

1.5.2 Python

Le choix de Python comme langage de programmation principal pour le développement du projet a été imposé par son utilisation préalable au sein de l'outil Sen2Chain. Python, en raison de sa simplicité, de sa flexibilité et de la richesse de ses bibliothèques spécifiques, a permis une écriture de code efficace, mettant l'accent sur la lisibilité et la maintenance. L'accès aux bibliothèques spécialisées de Python a également accéléré le développement en fournissant des solutions prêtes à l'emploi pour des tâches spécifiques. Ainsi, l'adoption stratégique de Python a offert un équilibre optimal entre puissance et facilité d'utilisation pour le développement du projet.

1.5.3 X2Go

L'environnement de travail comprenait plusieurs serveurs dédiés, chacun jouant un rôle spécifique dans le processus. Pour faciliter l'accès et la gestion, nous avons utilisé X2Go, un système de serveur de sessions graphiques qui permettait de se connecter à des machines virtuelles. Ces

dernières, configurées pour répondre aux besoins spécifiques du projet, constituaient un élément essentiel de notre infrastructure.

1.5.4 Utilisation de différents serveurs

L'environnement de travail de SEAS-OI se compose de plusieurs serveurs dédiés, comprenant 8 serveurs physiques, chacun associé à une partie spécifique du stockage. La virtualisation est assurée par VMware et Proxmox, ce dernier étant un logiciel libre. Trois machines physiques utilisent VMware, tandis que cinq autres machines exploitent Proxmox.

La répartition des machines virtuelles sur des serveurs physiques distincts permet d'optimiser la puissance de traitement. Le balancement des serveurs virtuels est géré manuellement en fonction de la charge des serveurs physiques.

À titre d'exemple, la machine "Sentinel" dispose de 64 CPU, 100 gigaoctets de RAM et 100To d'espace disque au total. Cette machine possède des données sur un SAN rapide et partage des fichiers via le protocole NFS.

Le stockage utilise d'une part des NAS(Network Attached Storage) QNAP dans un réseau IP dédié et d'autre part, un réseau de stockage SAN(Storage Area Network) afin d'améliorer les performances et l'accès aux données. Cette infrastructure permet une migration rapide des machines virtuelles entre les serveurs physiques et les pools de stockages.

La centralisation du de stockage optimise les coûts, les ressources et les capacités. Elle permet d'ajuster plus facilement les capacités et les performances du système afin de répondre à l'évolution constante des besoins de l'infrastructure. Elle simplifie également la gestion et réduit le risque de panne système.

Les interfaces de VMware et Proxmox sont utilisées pour superviser l'ensemble du système. Chaque serveur a une vue spécifique sur le stockage. L'accès aux espaces disques est configuré de manière sélective via l'administration des baies de stockage qui gère les autorisations pour chaque serveur.

Des sauvegardes journalières et hebdomadaires sont effectuées sur disques et sur bandes. Elles représentent un total d'environ 150 téraoctets, permettant de garantir la sécurité de l'ensemble des données. Un serveur est totalement dédié à la sauvegarde.

Un schéma de cette infrastructure est disponible sur la figure 15 de l'annexe .

Chapitre 2

Ma mission

Les données satellitaires sont une ressource précieuse pour de nombreuses applications, offrant des informations précises sur notre planète depuis l'espace. Ces données proviennent principalement de satellites optiques et radar, offrant une résolution allant de haute à très haute (échelle submétrique à décamétrique). Elles sont rendues accessibles grâce à divers dispositifs tels que les pôles KALIDEOS et DINAMIS de l'infrastructure de recherche Data Terra, et de la plateforme PEPS du CNES. Récemment, l'engagement dans des programmes de recherche a permis la mise en place à la station SEAS-OI, de chaînes de traitement opérationnelles, flexibles et spécialisées dans l'analyse d'images satellites Sentinel-1 et Sentinel-2 du programme européen Copernicus¹. L'objectif principal de mon stage était de contribuer de la chaîne de traitement des données SENTINEL-2.

2.1 L'outil Sen2Chain

L'outil Sen2Chain a été initié pour le projet S2-Malaria, il permet l'automatisation l'acquisition et le traitement d'images Sentinel-2 pour le calcul de séries temporelles d'indices. Aujourd'hui, il est principalement coordonné par Christophe Révillon, Pascal Mouquet et Vincent Herbreteau, avec le soutien de Didier Bouche (Université de La Réunion). Le projet a bénéficié de l'infrastructure SEAS-OI. Les premiers scripts ont été développés lors d'un stage d'ingénierie ENSG par Charlotte Wolff (1) en 2017, suivi d'une seconde phase de codage plus formelle par Jérémy Commins de 2017 à 2018. Les développements se sont poursuivis depuis septembre 2018, dans le cadre du projet FEDER RenovRisk-Impact (2), et Thomas Germain a rejoint l'équipe de développement en septembre 2022.

2.1.1 Données de Sentinel-2

SENTINEL-2 est une composante importante du programme européen Copernicus. La Commission Européenne fixe ses objectifs et assure la responsabilité financière de Copernicus, tandis que l'ESA est chargée de sa mise en application en collaboration avec les instances européennes et les agences nationales. L'organisation européenne pour les satellites météorologiques (EU-METSAT) et le centre européen de prévision météorologique (ECMWF) sont aussi partie intégrante de Copernicus.

L'ESA est responsable des opérations du segment sol des missions SENTINEL-1,2,3 et du satellite SENTINEL-5P, qui collectent chaque jour des milliers d'images. Ces données sont disponibles de façon ouverte et gratuite aux utilisateurs, avec comme objectif : mieux comprendre notre planète et améliorer la qualité de vie des citoyens européens. Les produits offrent une diversité de paramètres décrivant la végétation, le bilan énergétique et le cycle de l'eau. Les données recueillies permettent d'extraire des paramètres liés à la biogéophysique, tels que l'occupation des sols et les changements associés. Ces informations permettent une surveillance systématique et une cartographie des propriétés physiques globales influencées par l'activité humaine, ainsi que des variations saisonnières de la couverture végétale à l'échelle mondiale.

La mission SENTINEL-2 du programme Copernicus se compose de deux satellites identiques, SENTINEL-2A lancé en juin 2015 et SENTINEL-2B lancé en 2017. La figure figure 2.1 présente une vue schématique de ces engins.

Les deux satellites SENTINEL-2 opèrent simultanément, en phase à 180° l'un par rapport à l'autre, sur une orbite héliosynchrone à une altitude moyenne de 786 km, permettant une période de revisite de 5 jour. La position de chaque satellite SENTINEL-2 sur son orbite est mesurée par

1. Plus d'informations sur Copernicus : <https://omcs.atlassian.net/wiki/spaces/CSWK/pages/407404661/Copernicus+Programme>

un récepteur GNSS (Global Navigation Satellite System). La précision orbitale est maintenue par un système de propulsion dédié. Un mécanisme d'obturation empêche l'instrument d'être éclairé directement par le soleil en orbite. Ce mécanisme fonctionne aussi comme un dispositif d'étalonnage en recueillant la lumière du soleil après réflexion par un diffuseur.

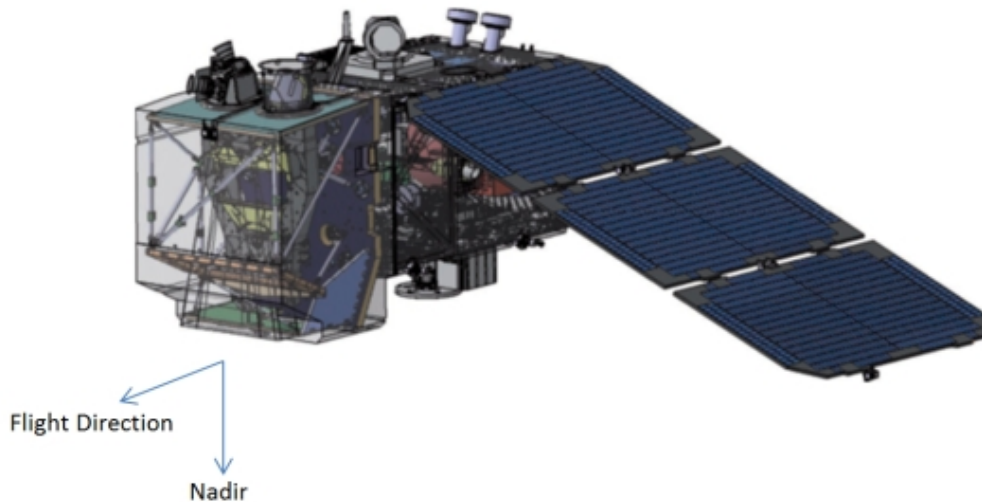


FIGURE 2.1 – Vue schématique de l'engin Sentinel-2 [Credits : EADS Astrium]

Dans la pratique, les données utilisables sont souvent plus espacées que la période de revisite, dû au fait que certaines images sont inexploitable à cause de la couverture nuageuse.

Chaque satellite transporte un instrument multispectral MSI², construit par Airbus Defense and Space en France, qui permet d'acquérir les images. L'instrument est un ensemble de capteurs optiques passifs et mesure la radiance réfléchi de la Terre en 13 bandes spectrales dans le domaine du visible, du proche infrarouge et de l'infrarouge à ondes courtes, à différentes résolutions au sol. Les satellites SENTINEL-2 offrent une résolution spatiale d'environ 10 mètres (haute résolution spatiale). Cela permet une surveillance à haute résolution l'évolution de la végétation, de l'occupation des sols et l'impact du réchauffement climatique avec une haute fréquence d'observations. SENTINEL-2 dispose d'un champ de vision de 290 km lors de la capture de ses images.

La durée de vie du satellite est de 7,25 ans, ce qui comprend une phase de mise en service en orbite de 3 mois. Des batteries et des ergols ont été fournis pour tenir 12 ans d'exploitation, y compris les manœuvres de désorbitation en fin de vie.

Ayant exploré la mission Sentinel-2 et son rôle crucial dans la collecte de données satellitaires à haute résolution, approfondissons maintenant le processus mis en place à la station SEAS-OI pour la récupération des images et la génération des produits qui en résulte. Cette phase implique des étapes essentielles telles que la segmentation des images en tuiles, la capture et le traitement des données, ainsi que l'application de différents niveaux de traitement. Chacun de ces éléments joue un rôle déterminant dans la transformation des données brutes en informations exploitables, ouvrant ainsi la voie à une compréhension approfondie et à des applications diverses dans les domaines de la cartographie, de la surveillance environnementale et bien d'autres.

2. Plus d'informations sur l'instrument MSI : <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/instrument-payload>

Tuiles

La Terre est subdivisée sur un ensemble prédéfini de tuiles, défini dans la projection UTM / WGS84. La projection UTM (Universal Transverse Mercator) est un système de coordonnées géographique qui est basé sur la projection transversale de Mercator qui est très utilisé dans les systèmes de cartographie nationaux et internationaux du monde entier. Dans UTM-WGS84, l'UTM est projeté sur l'ellipsoïde WGS84 (World Geodetic System) qui est le système de référence utilisé par le Global Positioning System (GPS). Chaque tuile utilise un pas de 100km mais a une surface fixe de 110x110 km² afin d'offrir un grand chevauchement avec le voisinage. Il y a un chevauchement d'environ 5 km entre les tuiles adjacentes dans la même zone UTM et sur la bordure de la zone UTM, il y a un chevauchement plus important avec les tuiles de la zone UTM adjacente.

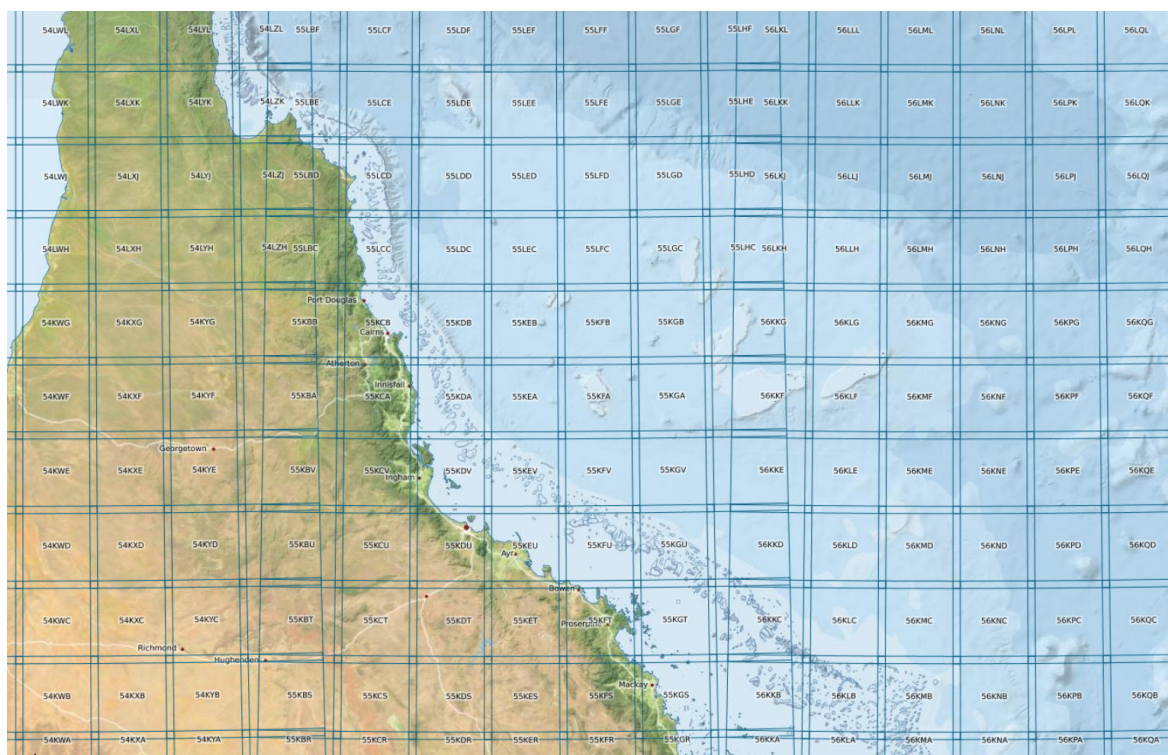


FIGURE 2.2 – Aperçu du découpage [Crédits : eAtlasInstitut]

Chaque tuile est identifiée par un ID unique. L'ID suit un schéma spécifique permettant de localiser toutes les images associées à une tuile donnée. Le monde est divisé en 60 zones de 8 degrés de largeur en longitude, numérotées de 1 à 60 en se déplaçant vers l'est, avec la zone 1 située sur l'océan Pacifique. Dans une désignation de tuile telle que 11SUJ, les 2 premiers chiffres et la lettre 1 suivante correspondent respectivement à la zone UTM et à la bande de latitude. Chaque zone est subdivisée en bandes de 6 degrés de latitude, représentées par des lettres qui progressent du Sud vers le Nord. Le quatrième caractère, d'est en ouest, et le cinquième caractère, du sud au nord, identifient ces subdivisions.

Prise et bande de donnée

Les satellites SENTINEL-2 acquièrent systématiquement des données sur les zones terrestres et côtières dans une bande de latitude allant de 56° Sud (Isla Hornos, Cap Horn, Amérique du Sud) à 83° Nord (au-dessus du Groenland). La longueur maximale d'une prise de données d'imagerie est de 15 000 km. De plus, les satellites SENTINEL-2 acquièrent des observations supplémentaires au-dessus de sites d'étalonnage spécifiques, tels que Dome-C en Antarctique.

La combinaison de la grande largeur de fauchée (290 km), des bandes spectrales multiples et de la haute résolution spatiale, associée à l'exigence d'acquisition de données globales avec une fréquence de revisite élevée génère quotidiennement 1,6 TByte de données d'images brutes compressées à partir de la constellation de deux satellites.

Les produits contiennent des tuiles provenant d'une seule prise de donnée. Dans une prise de donnée, une portion d'image transmise à une station donnée lors d'un passage est appelée bande de donnée. Une prise de donnée est présentée à l'intérieur d'un produit comme un ensemble d'une ou plusieurs bandes de données qui correspondent à des segments d'acquisition transmis à plusieurs stations de réception au sol.

Niveaux de traitement

Les données satellitaires sont proposées à plusieurs niveaux de traitement. Les traitements de niveau 0, 1A et 1B ne sont généralement pas accessibles aux utilisateurs.

Le produit de niveau 0 est constitué de données d'images brutes. Elles sont initialement traitées pour protéger l'intégrité des données acquises, ce traitement s'effectue parallèlement à la réception des données.

Le produit 1A est obtenu en décompressant les données brutes de l'image de niveau 0. Un modèle géométrique est développé, permettant de localiser n'importe quel pixel de l'image.

Le produit 1B fournit des images corrigées radiométriquement.

Les produits de niveau 1C sont des images corrigées radiométriquement et géométriquement. Ces corrections comprennent l'orthorectification et le recalage spatial sur le système de référence global. Les données sont disponibles au format JPEG2000 avec compression sans perte, accompagnées des métadonnées, dans un conteneur de fichiers SAFE. Les produits sont constitués de tuiles, ils résultent de l'utilisation d'un modèle numérique d'élévation pour projeter l'image dans une géométrie cartographique, corrigeant les distorsions géométriques du sol. Les mesures radiométriques par pixel sont fournies en réflectance au sommet de l'atmosphère, avec des paramètres pour les convertir en radiances. Les produits sont rééchantillonnés à une distance d'échantillonnage au sol constante de 10, 20 et 60 m en fonction de la résolution native des bandes spectrales, et les coordonnées des pixels se réfèrent au coin supérieur gauche.

Les produits de niveau 2A ne sont actuellement pas générés systématiquement, mais peuvent être générés par les utilisateurs eux-mêmes. Le produit 2A fournit des images de réflectance de surface corrigées de l'atmosphère, dérivées des produits 1C associés. Cette correction atmosphérique permet d'éliminer les effets des composants atmosphériques sur les mesures radiométriques, fournissant ainsi des données plus précises et cohérentes de la réflectance de la surface terrestre.

Il convient de noter que cette situation est en train de changer, et ces produits de niveau 2A sont de plus en plus disponibles en temps réel, marquant une évolution positive dans l'accessibilité et la rapidité de disponibilité de ces données.

2.1.2 La chaîne de traitement Sen2Chain

Sen2Chain, écrit en Python 3, constitue une solution complète pour la gestion et le traitement d'images SENTINEL-2, récupérées depuis les catalogues de l'ESA (SciHub) et du CNES (PEPS). Il simplifie les opérations de téléchargement, de correction et d'analyse radiométrique, tout en optimisant l'utilisation des ressources grâce à la parallélisation. Le code est librement accessible sur Framagit pour d'autres utilisateurs et développements. L'intégration de la programmation orientée objet dans Sen2Chain permet d'améliorer la modularité, de faciliter la

maintenance, d'encourager la réutilisation du code et de favoriser la collaboration au sein de l'équipe de développement. Cette approche simplifie le processus de développement de l'outil tout en le rendant plus adaptable et extensible.

Récupération des images

La récupération des produits de niveau 1C par Sen2Chain se fait grâce à EODAG³ (Earth Observation Data Access Gateway). EODAG, se présente comme un outil en ligne de commande et une bibliothèque Python. L'outil a été développé par CS GROUP et est disponible sous la licence open source Apache License. Son objectif est de faciliter la recherche et le téléchargement d'images obtenues par télédétection, tout en offrant une API unifiée pour accéder aux données, quel que soit le fournisseur. Cet outil permet un accès aisé à une variété de produits provenant de plus de 10 fournisseurs, englobant plus de 50 types de produits distincts tels que Sentinel 1, Sentinel 2, Sentinel 3, Landsat, et bien d'autres. Sa documentation complète est également facilement accessible en ligne, facilitant ainsi son utilisation et son intégration.

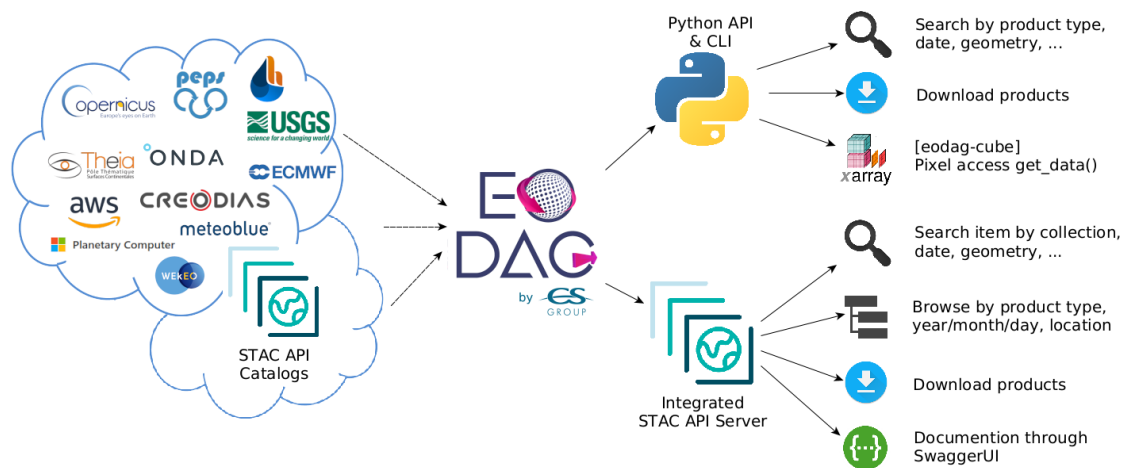


FIGURE 2.3 – Aperçu d'EODAG [Crédits : CS GROUP]

Une fois récupérés, les produits 1C sont traités par Sen2Chain à l'aide de l'algorithme Sen2Cor de l'ESA. Sen2Cor est un processeur pour la génération et le formatage de produits SENTINEL-2 de niveau 2A. L'algorithme du processeur Sen2Cor est une combinaison de techniques de pointe pour effectuer des corrections atmosphériques qui ont été adaptées à l'environnement SENTINEL-2.

Production des cloudmasks

Les produits de niveau 2A sont alors traités afin de produire le cloudmask correspondant. La détection précise des nuages dans les images de télédétection optique revêt une importance cruciale, car des nuages non détectés peuvent introduire des erreurs dans l'analyse des paramètres atmosphériques et de surface. Les nuages ont un impact significatif sur la transmission du rayonnement entre les capteurs satellitaires et les cibles de surface, entraînant une diminution de la capacité à identifier les caractéristiques terrestres. Cela compromet la précision de la récupération des paramètres de surface et peut fournir des données trompeuses sur les zones surveillées.

3. Plus d'informations : <https://eodag.readthedocs.io/en/stable/index.html>

Calcul des indices

Sen2Chain propose une gamme de 11 indices radiométriques pour l'analyse. Ces indices sont des mesures dérivées des valeurs radiométriques des différentes bandes spectrales des images satellitaires. Ils fournissent des informations ciblées sur la couverture terrestre, telles que la végétation, l'eau, et d'autres caractéristiques.

Le calcul des indices se fait à partir des différentes bandes spectrales des images L2A et des masques nuages. Ces indices sont essentiels pour des analyses spécialisées, telles que la surveillance des cultures, la détection des changements environnementaux, et l'évaluation des ressources naturelles. L'utilisation des masques nuage lors du calcul de ces indices garantit des résultats fiables en éliminant les zones nuageuses, préservant ainsi la précision des calculs radiométriques.

Extractions

L'outil Sen2Chain dispose de la possibilité de réaliser des extractions de séries temporelles à partir des indices spectraux disponibles dans la base de données. Les données extraites sont converties en un fichier CSV, facilitant ainsi la récupération de données structurées pour des analyses ultérieures. L'extraction se concentre sur les sites d'intérêt définis par l'utilisateur, qui peuvent être spécifiés sous forme de points ou de polygones au format shapefile. L'utilisateur a le pouvoir de choisir la période d'analyse, le type de masque nuage, et les indices à inclure. Une fonction de lissage optionnelle est également disponible pour corriger les valeurs aberrantes dans les séries temporelles, offrant ainsi une interpolation pour obtenir une valeur quotidienne.

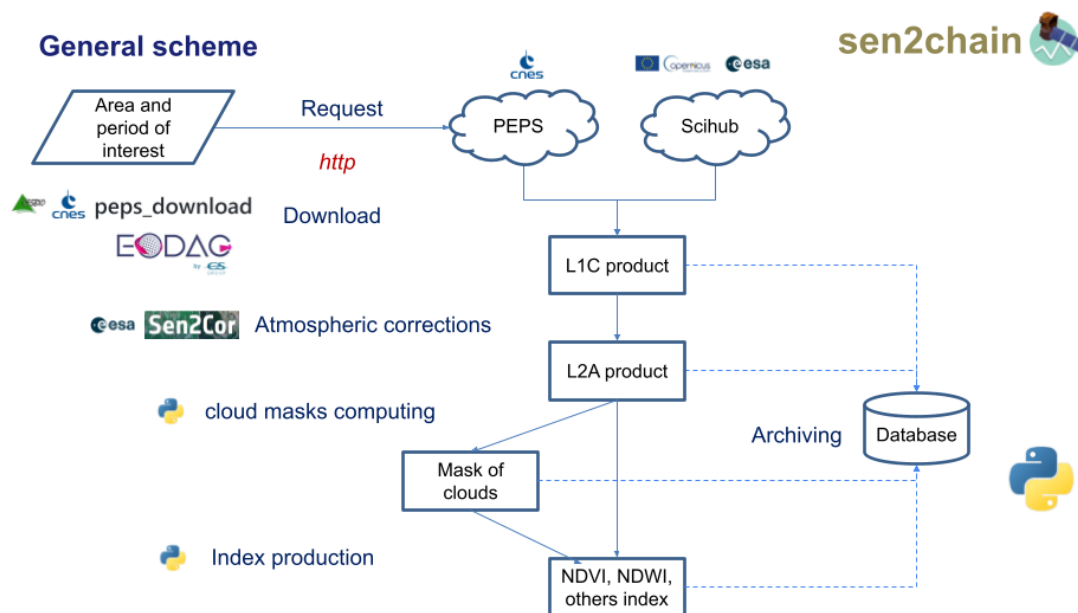


FIGURE 2.4 – Schéma général du fonctionnement de Sen2Chain

Arborescence de sortie

Sen2Chain organise ces données dans une arborescence spécifique permettant d'ordonner et de trouver facilement les différents produits.

Les produits de niveau 1C sont sauvegardés dans le répertoire L1C. Les produits de niveau 2A sont sauvegardés dans le répertoire L2A. À l'intérieur des dossiers L1C et L2A, un dossier

distinct est créé pour chaque tuile. À l'intérieur de chaque dossier de tuile, il y a les produits qui correspondent à cette tuile. Tous les produits, 1C ou 2A, suivent la convention de nommage des produits, facilitant ainsi la gestion et la localisation des produits associés à chaque tuile. Le nom de produit suit le format général suivant :

```
MMM_MSIXXX_YYYYMMDDHHMMSS_Nxxyy_R000_Txxxxx_<Product Discriminator>.SAFE
```

- MMM : Il s'agit de l'ID de mission, pouvant être S2A (SENTINEL-2A) ou S2B (SENTINEL-2B), indiquant le satellite utilisé.
- MSIXXX : Ce composant indique le niveau de produit, où MSIL1C désigne le produit de niveau 1C et MSIL2A désigne le produit de niveau 2A.
- YYYYMMDDHHMMSS : Cette partie représente l'heure de début de détection de la prise de données.
- Nxxyy : Il s'agit du numéro de ligne de base de traitement, qui sera explicité plus tard.
- R000 : C'est le numéro d'orbite relative, avec des valeurs allant de R001 à R143.
- Txxxxx : Indique le numéro de tuile.
- <Product Discriminator> : Ce champ de 15 caractères sert à distinguer différents produits destinés à l'utilisateur final provenant de la même prise de données. Selon l'instance, l'heure dans ce champ peut être antérieure ou légèrement postérieure à l'heure de détection de la prise de données.
- SAFE : Indique le format du produit, à savoir le Format d'Archive Standard pour l'Europe.

Ainsi, un exemple concret serait le nom de fichier :

```
S2A_MSIL1C_20170105T013442_N0204_R031_T53NMJ_20170105T013443.SAFE
```

Cette dénomination identifie un produit de niveau 1C acquis par le satellite SENTINEL-2A le 5 janvier 2017 à 01 :34 :42 (temps universel), sur la tuile 53NMJ lors de l'orbite relative 031, et traité avec la ligne de base de traitement 02.04.

Ensuite, tout les produits dérivés des produits de niveau 2A sont stockés dans un répertoire appelé "PRODUIT".

Une fois les masques nuage produits, un fichier raster ainsi qu'un fichier XML sont créés. Ils respectent la convention de nommage du produit L2A dont ils proviennent, et le nom du masque nuage est ajouté à la fin. Les produits sont, comme les produits 1C et 2A, répartis par tuiles. Un exemple de nom de produit est le suivant :

```
S2A_MSIL1C_20170105T013442_N0204_R031_T53NMJ_20170105T013443_CM004-CSH1-  
CMP1-TCI1-ITER1.jp2
```

De même pour les indices radiométriques, les produits suivent la convention de nommage du produit L2A masqué dont ils sont dérivés, et le nom de l'indice est ajouté à la fin. Les produits sont, comme les produits 1C et 2A, répartis par tuiles. Voici un exemple du nom d'un produit :

```
S2A_MSIL1C_20170105T013442_N0204_R031_T53NMJ_20170105T013443_NDVI_CM004-CSH1-  
CMP1-TCI1-ITER1.jp2
```

2.2 Objectifs de ma mission

Au cours de mon stage, la priorité était de participer activement au développement de l'outil Sen2chain, en collaboration avec les développeurs des chaînes de traitements de la station.

L'objectif était de proposer des améliorations, d'ajouter de nouvelles fonctionnalités et d'intégrer opérationnellement ces modifications aux outils existants. Les étapes du projet étaient les suivantes :

1. **Prise en main de l'environnement de travail** : Cela impliquait la familiarisation avec les outils de la station SEAS-OI dédiés au traitement des images Sentinel-2. Il a fallu installer et analyser les codes informatiques existants.
2. **Identification de points d'amélioration de la chaîne** : Parmi les axes d'évolution déjà identifiés figuraient le nettoyage et l'optimisation du code, la création d'une base de données pour une meilleure gestion des produits dans la base de données image de la station (actuellement > 50 To), et l'amélioration des extractions des séries temporelles d'indices spectraux.
3. **Élaboration d'algorithmes de traitement** : Il s'agissait de concevoir des algorithmes de traitement, de les implémenter, de coder ces solutions, et enfin de les intégrer dans la chaîne de traitement, tout en assurant une mise en production efficace.
4. **Rédaction de la documentation** : La documentation associée au projet devait être rédigée, et les codes devaient être mis en ligne sur la forge logicielle de l'UMR.
5. **Correction de bugs, maintenance, soutien aux utilisateurs** : Enfin, le stage incluait également la résolution de bugs éventuels, la maintenance des solutions mises en place, et le soutien aux utilisateurs.

Ce processus rigoureux visait à garantir le succès du développement de Sen2chain, en assurant une collaboration efficace et en répondant aux besoins opérationnels de la station SEAS-OI.

Un cahier des charges détaillé a été élaboré pour définir précisément les spécifications du projet, les exigences techniques, ainsi que les contraintes à prendre en compte. Parallèlement, un planning méthodique a été établi, guidant les différentes étapes du projet, les échéances à respecter, et les jalons cruciaux. Dans la partie suivante, je décrirai de manière détaillée le déroulement du projet, mettant en lumière les étapes clés et les résultats obtenus.

2.2.1 Mise en place du cahier des charges

Au fil de discussions régulières avec mon tuteur et à mesure que les objectifs du stage prenaient une forme plus précise, la formalisation du cahier des charges a été entreprise. Ces échanges ont constitué une opportunité cruciale pour clarifier les attentes, identifier les contraintes et définir les étapes clés du projet. L'évolution du cahier des charges, au rythme des progrès du code et des dialogues continus, témoigne de l'affinement constant des objectifs et de la convergence vers une feuille de route cohérente pour guider la réalisation des tâches à venir.

2.3 Planning

Initialement, lors du début du projet, un chronogramme provisoire a été établi, accompagné d'un cahier des charges détaillé et de clauses techniques spécifiques pour chaque tâche prévue. Cependant, au fil de l'avancement de mon stage, certaines évolutions ont été nécessaires, notamment en ce qui concerne l'évaluation du temps nécessaire à chaque tâche et l'émergence de nouvelles idées qui ont entraîné des ajustements dans la planification. La complexité du projet a rendu la planification des tâches particulièrement délicate, car les nouvelles idées ont souvent conduit à des modifications de tâches, et l'impact de ces changements sur les échelles de temps initialement envisagées s'est révélé difficile à quantifier. Comme le montre l'extrait

du chronogramme initial présenté dans la figure 2.5, le stage était composé de diverses tâches variées, ce qui constituait l'une des caractéristiques clés du projet.

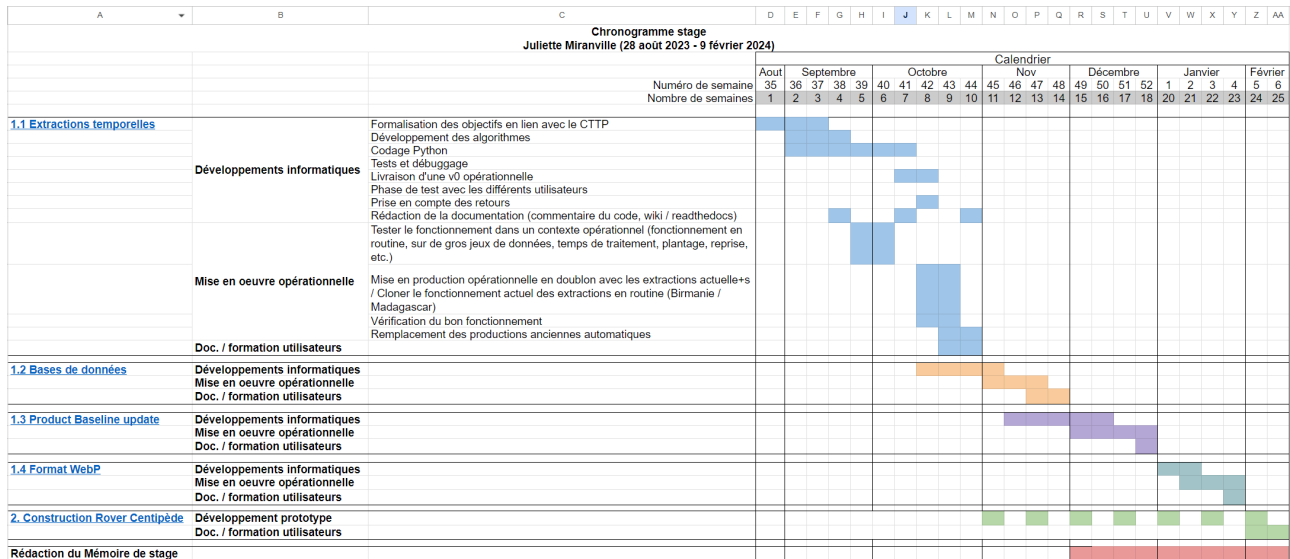


FIGURE 2.5 – Extrait du chronogramme initial

Finalement, le chronogramme a évolué pour refléter les ajustements nécessaires, comme le montre la figure 2.6, représentant le chronogramme final du stage. Cette flexibilité dans la planification a permis de s'adapter aux challenges rencontrés et d'intégrer de manière cohérente les nouvelles idées qui ont enrichi le développement du projet. Cette expérience a souligné

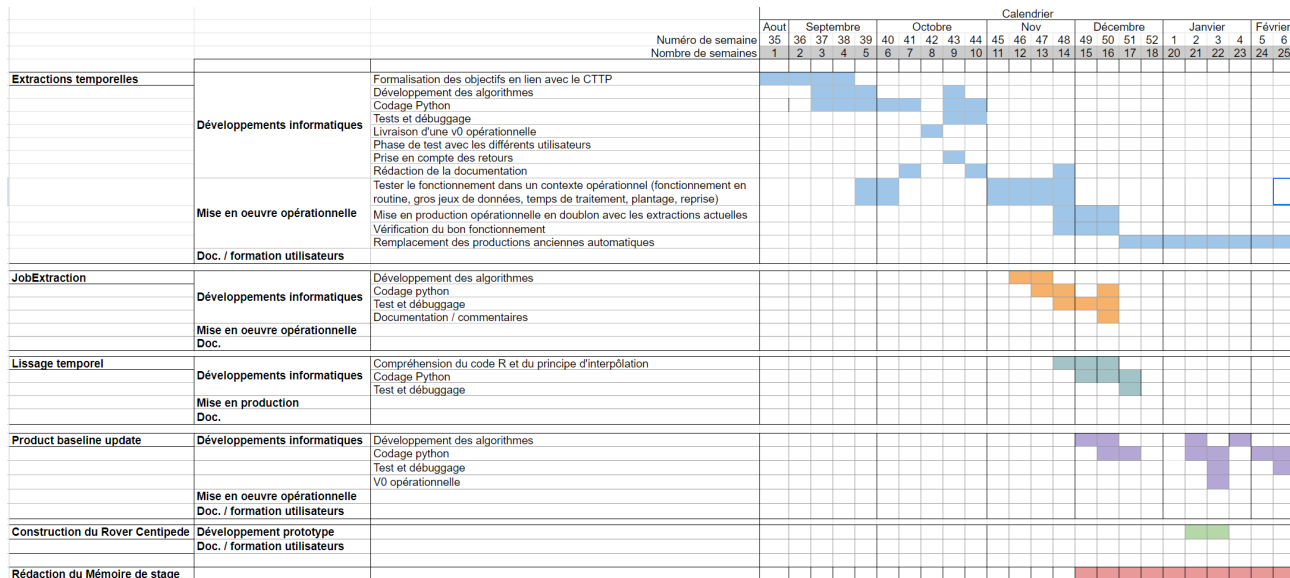


FIGURE 2.6 – Chronogramme final

l'importance de la flexibilité dans la gestion de projet, en particulier dans un contexte où l'innovation et l'émergence d'idées nouvelles peuvent influencer significativement la direction du travail. La diversité des tâches entreprises a également constitué un atout majeur du stage, offrant une expérience riche et variée dans le cadre de la réalisation du projet.

Chapitre 3

Déroulement du projet

Dans cette section, nous examinerons les ajustements techniques effectués au cours de notre projet, centré sur l'amélioration de Sen2Chain, un outil de traitement des produits SENTINEL-2. Ces modifications ont joué un rôle essentiel dans l'évolution de Sen2Chain.

3.1 Appropriation du sujet

L'initialisation du projet a nécessité un temps d'appropriation de l'outil Sen2Chain. La compréhension du domaine a impliqué l'assimilation de nombreux concepts et la familiarisation avec un vocabulaire spécifique. Cette phase a été appuyée par une exploration de diverses ressources disponibles, contribuant ainsi à établir une base solide pour la suite du travail.

Parallèlement, la compréhension du code représentait un enjeu crucial. J'ai initié ce processus en suivant la documentation d'installation et de configuration de la chaîne, ce qui m'a offert l'opportunité de formuler des observations critiques concernant la documentation existante. Ce retour a été explicitement sollicité dans le but d'améliorer les documents en question. Puis, j'ai ensuite effectué différents tests de l'outil, comprenant le téléchargement de données, la production de masque et indices et l'extraction de données afin de comprendre globalement le fonctionnement de la chaîne. En concentrant mes efforts initiaux sur les points que je devais modifier, j'ai pu progressivement appréhender l'ensemble de la structure de l'outil Sen2Chain. La documentation a été un outil précieux, fournissant des indications claires sur le fonctionnement du code. Mes premiers jours de stage ont ainsi été dédiés à manipuler l'outil, renforçant ainsi ma familiarité avec le projet dans son ensemble.

Le code faisait appel à plusieurs bibliothèques Python, nécessitant ainsi une familiarisation avec celles-ci. Parmi ces bibliothèques, j'ai particulièrement approfondi mes connaissances sur la bibliothèque Pandas, étant donné son usage fréquent dans le cadre de manipulations de données. Les DataFrames, largement employés dans le code, ont constitué un aspect essentiel de cette bibliothèque, m'amenant à acquérir une expertise spécifique pour traiter et analyser efficacement les données au sein du projet. Ce processus m'a permis d'optimiser ma compréhension des fonctionnalités offertes par Pandas et de les appliquer de manière pertinente dans le contexte du projet.

3.2 Mise à jour de l'outil d'extraction des séries temporelles

3.2.1 La classe TimeSerie

La classe `TimeSerie` permet d'effectuer les extractions des séries temporelles. Lors de la création d'une instance de cette classe, l'utilisateur spécifie les sites d'intérêt, pouvant être définis sous forme de points ou de polygones au format shapefile.

Un shapefile, est un format de fichier utilisé dans les systèmes d'informations géographiques (SIG). Il a une extension classique en SHP et est toujours accompagné d'autres fichiers portant le même nom mais avec les extensions suivantes :

- `.dbf` : un fichier qui contient les données attributaires relatives aux objets contenus dans le shapefile.
- `.shx` : un fichier qui stocke l'index de la géométrie.
- `.shp` : le format de la forme, représentant la géométrie de l'entité elle-même.
- `.cpg` : un fichier facultatif utilisé pour spécifier la page de code des fichiers `.dbf` (content-type : `text/plain` ou `x-gis/x-shapefile`).

- `.prj` : contient des informations sur le système de coordonnées, utilisant le format WKT (well-known text).

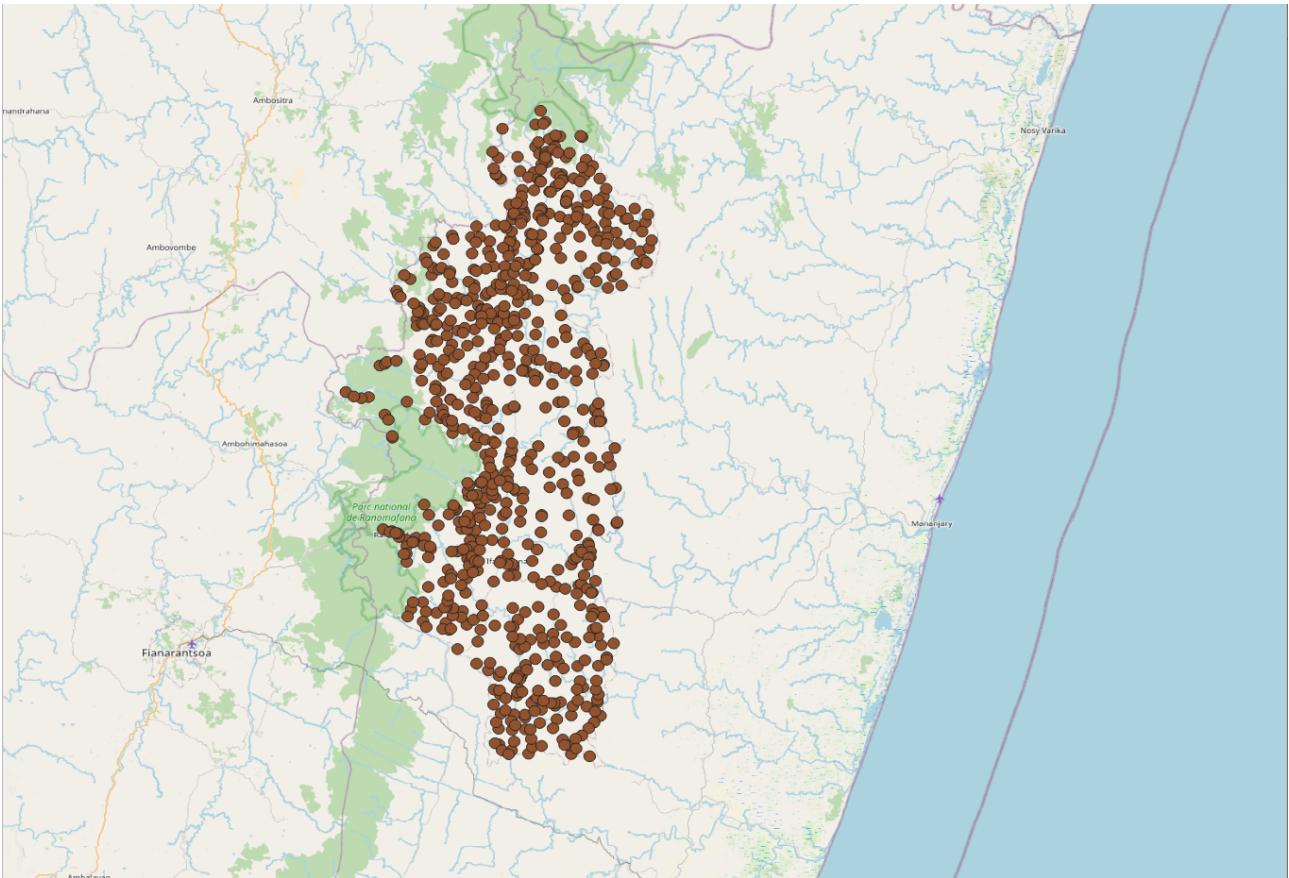


FIGURE 3.1 – Exemple de shapefile représentant des villages sur Madagascar

L'utilisateur délimite ses zones d'intérêt en utilisant ces fichiers. Par exemple, sur l'image figure 3.1, les zones d'intérêt sont indiquées en marron. Ces zones peuvent être représentées sous forme de points ou de polygones.

De plus, l'utilisateur renseigne aussi : une liste d'indices radiométriques disponibles via Sen2Chain, un masque nuage et ses paramètres en fonction des besoins spécifiques de l'analyse ainsi que la dates minimale et maximale délimitant la période d'intérêt. Ces informations étant introduites lors de la création d'une instance de la classe `TimeSerie`, cela offre une personnalisation complète des extractions temporelles selon les critères et les préférences de l'utilisateur.

L'utilisation des ressources nécessaire aux extractions des séries temporelles est optimisé grâce à la parallélisation.

Une classe `Produit` existe afin de faciliter l'accès à aux produits de niveau 1C, 2A, les cloudmasks et les indices.

Pour assurer une compréhension approfondie du fonctionnement de cette classe, une phase de test du code a été effectuée. Ensuite, les ajustements nécessaires ont été apportés afin de répondre aux objectifs définis par le cahier des charges.

3.2.2 Objectifs

Le fonctionnement de l'outil d'extraction des séries temporelles présentait plusieurs contraintes significatives qui limitaient sa praticité en environnement opérationnel. Tout d'abord, le temps

nécessaire pour effectuer l'extraction est excessivement long, notamment lorsque le shapefile d'entrée est volumineux, comportant de nombreuses entités et de vastes surfaces à extraire. De plus, il n'existe pas de mécanisme de sauvegarde des données en cours de traitement, ce qui signifiait que tout dysfonctionnement du code entraînait la perte de nombreuses heures de calculs, que ce soit en raison d'erreurs dans les codes, de bugs dans les données, ou même de plantages du serveur utilisé simultanément pour d'autres tâches de calcul parfois intensives.

Un autre problème majeur résidait dans l'absence de gestion intelligente des produits déjà extraits. Lorsqu'une extraction était lancée, le système ne vérifiait pas ce qui avait déjà été extrait, mais relançait l'ensemble du processus depuis le début, ce qui impliquait de traiter à nouveau l'ensemble de la série temporelle, parfois composée de plusieurs centaines de dates. En outre, le format de sortie des données se présentait sous forme d'un unique fichier CSV, devenant de plus en plus volumineux avec le temps. Cette accumulation de données pouvait rendre la manipulation et l'analyse des informations extraites de plus en plus complexes, impactant la lisibilité des résultats. Aussi, la volumineuse taille du fichier de résultat pouvait restreindre sa facilité d'exploitation et de diffusion via le catalogue en ligne de SEAS-OI, un canal privilégié pour les échanges avec les partenaires étrangers, en particulier dans des régions telles que Madagascar, où les connexions aux réseaux et les débits de téléchargement sont limités.

Dans le but d'améliorer ces aspects du système et de le rendre plus adapté à une utilisation opérationnelle efficace, les objectifs du projet ont été définis comme suit :

- **Simplification de l'extraction de données** : rendre l'extraction des données la plus simple possible, en limitant la nécessité d'écrire des lignes de code complexe.
- **Garantie de la sauvegarde des données** : Assurer la sécurité des données par des enregistrements réguliers des résultats d'extraction.
- **Automatisation intelligente** : Mettre en place une automatisation intelligente pour éviter le retraitement inutile de produits déjà extraits. Gérer efficacement les ajouts de nouveaux produits, y compris dans la période temporelle.

Par la suite, une exposition des tâches accomplies en réponse à ces objectifs sera présentée.

3.2.3 Arborescence d'extraction

Une arborescence d'extraction visant à regrouper de manière organisée les produits résultant du processus d'extraction a été mise en place. Cette structure facilite grandement la récupération des extractions de manière efficace et méthodique.



FIGURE 3.2 – Vue de l'arborescence d'extraction

Chaque shapefile d'extraction dispose maintenant d'un dossier unique grâce à l'utilisation de la somme de contrôle (cf. partie 3.2.4).

Comme on peut l'observer sur la figure 3.2, plusieurs sous-dossiers ont été créés pour gérer les données de manière optimale.

Un sous-dossier "**shp_files**" qui contient une copie du shapefile d'extraction, offrant ainsi une référence pour les processus de vérifications ultérieurs. Le fait d'avoir une copie de ces fichiers

dans l'arborescence d'extraction sert de référence pour les processus de vérifications ultérieurs.

Un autre sous-dossier, nommé "**raw_data**" contient des fichiers individuels pour chaque produit extrait, organisés de manière claire et suivant la convention de nommage des produits. (cf. section 3.2.5)

Un dernier dossier, appelé "**compiled_data**" a été instauré pour regrouper les données compilées résultant du processus d'extraction. Cette structuration met en lumière la distinction entre les données brutes et celles qui ont été compilées, simplifiant ainsi la navigation et la gestion des informations extraites. Les données compilées seront celles à disposition de l'utilisateur. (cf. section 3.2.6)

Cette arborescence d'extraction offre non seulement une solution pratique pour retrouver rapidement les extractions, mais elle facilite également les vérifications, les analyses, et la gestion globale des données extraites, contribuant ainsi à une conduite efficace et ordonnée du projet.

3.2.4 Mise en place d'une somme de contrôle

Afin de permettre une reprise intelligente des extractions de séries temporelles, il était crucial de disposer d'une méthode fiable pour s'appuyer sur les données préexistantes. Il est en effet impératif de garantir une correspondance précise entre les données et le shapefile. Dans cette perspective, l'implémentation d'une somme de contrôle a été instaurée.

Une somme de contrôle, également connue sous le nom de checksum, est une méthode cryptographique utilisée pour vérifier l'intégrité des données. Elle génère une valeur numérique unique, appelée empreinte ou somme de contrôle, à partir d'un ensemble de données. Cette empreinte agit comme une signature digitale, permettant de détecter toute altération accidentelle ou malveillante des données. L'utilisation de sommes de contrôle est essentielle dans divers domaines, y compris la gestion de fichiers, la sécurité informatique et la transmission de données, afin de garantir la fiabilité des informations manipulées.

Lors de la création d'une instance de la classe TimeSerie, une méthode est déclenchée pour calculer la somme de contrôle des fichiers constituant le shapefile. Si le shapefile a déjà été utilisé dans le cadre d'une extraction de série temporelle, la somme de contrôle correspondante est présente dans l'arborescence. L'utilisateur est alors informé de cette situation et est notifié du chemin absolu où les données sont stockées. À ce stade, l'extraction peut être reprise de manière intelligente, un processus détaillé dans la section 3.2.5.

La vérification de ces empreintes numériques assure que les shapefiles qui sont utilisés en routine restent intacts au fil du temps, offrant ainsi une assurance supplémentaire contre toute altération des données. Cette approche permet donc de s'appuyer avec certitude sur les données existantes permettant la reprise intelligente des extractions et évitant la nécessité de faire une extraction complète à chaque utilisation.

Les noms des dossiers d'extraction sont générés de manière unique en concaténant le nom du shapefile avec sa somme de contrôle pour former le nom du dossier de sortie. Cela peut-être observer sur la figure 3.3.

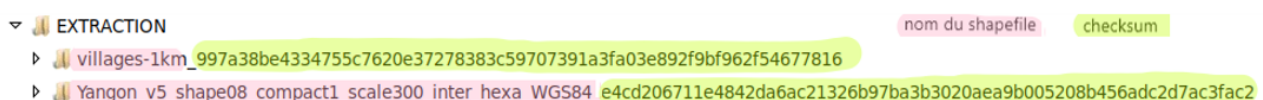


FIGURE 3.3 – Vue des dossiers d'extractions

Le choix du SHA-256 pour la mise en place d'une somme de contrôle revêt une importance cruciale dans le contexte de l'intégrité des shapefiles. Le SHA-256 (Secure Hash Algorithm 256 bits) est un algorithme de hachage cryptographique réputé pour sa robustesse et sa sécurité. En produisant une empreinte numérique de 256 bits, soit 32 octets ou 64 caractères hexadécimaux, cette fonction de hachage offre une représentation unique et difficilement falsifiable de chaque ensemble de données.

La longueur de la sortie du SHA-256 contribue à réduire considérablement les risques de collisions, où deux ensembles de données différents généreraient la même empreinte. Cette caractéristique renforce la sécurité du processus de vérification d'intégrité. De plus, la nature unique de chaque empreinte facilite la création d'arborescences distinctes pour chaque shapefile, simplifiant ainsi la gestion des données.

3.2.5 Données brutes

L'extraction des données des produits d'indices suit un processus défini. L'utilisateur commence par spécifier des zones d'intérêt sous la forme de polygones ou de points. Pour chaque indice demandé par l'utilisateur, le programme itère à travers toutes les tuiles correspondantes, c'est-à-dire celles englobant les polygones spécifiés. À partir de là, pour chaque tuile, le programme analyse les produits d'indices associés aux dates et cloudmask définis par l'utilisateur. Il extrait les informations pertinentes pour les polygones spécifiés.

Sauvegarde en cours d'extraction

Initialement, dans le processus, toutes les données extraites pour chaque shapefile étaient stockées dans des fichiers CSV volumineux, étant donné que les périodes temporelles d'extraction étaient généralement étendues, devenant de plus en plus volumineux avec le temps. Un fichier CSV distinct était créé pour chaque indice spécifié par l'utilisateur, et la dénomination suivait la structure `nomdushapefile_indice.csv`. Chaque CSV contient une ligne pour chaque polygone à chaque date disponible sur l'ensemble de la série temporelle.

Cependant, ce modèle présentait des inconvénients, particulièrement en cas d'échec pendant l'extraction, nécessitant de recommencer l'ensemble du processus car la sauvegarde des données par l'écriture du CSV sur le disque ne s'effectuait qu'à la toute fin de l'extraction. De plus, l'accumulation de données pouvait compliquer la manipulation et l'analyse des informations extraites, impactant la lisibilité des résultats. Afin d'améliorer cette méthode, une fonctionnalité de sauvegarde pendant l'extraction a été intégrée.

Dans un premier temps, le code permet la réalisation des extractions de manière séquentielle. La sauvegarde peut alors s'effectuer à chaque produit extrait, garantissant ainsi la persistance des données.

L'introduction du multiprocessing dans le code, indispensable pour traiter de manière optimale très grand volume de données, a apporté une complexité supplémentaire. Plusieurs produits de la même tuile étaient traités simultanément, nécessitant l'utilisation d'un dictionnaire partagé pour stocker les données. La sauvegarde doit alors être gérée de manière à s'assurer que toutes les données pour une tuile étaient extraites avant de procéder à la sauvegarde. La sauvegarde par l'écriture des CSV sur le disque, ne se fait donc qu'après le traitement complet d'une tuile.

Reprise intelligente de l'extraction

Une autre problématique était la reprise d'extraction ou la continuation d'extractions inachevées ou incomplète après le téléchargement de nouvelles données. Initialement, lorsqu'une

extraction était lancée, le système ne pouvait pas vérifier ce qui avait déjà été extrait, il devait alors relancer l'ensemble du processus. Cela entraînait une utilisation intensive des ressources et une augmentation significative du temps d'extraction. Pour éviter de traiter à nouveau les mêmes données, une vérification basée sur les dates a donc été mise en place dans un premier temps. Initialement, l'idée était de reprendre l'extraction à partir de la dernière date traitée. Cependant, cette approche s'est avérée insuffisante, car de nouveaux produits pouvaient être introduits avec des dates antérieures à la dernière date traitée et risquaient donc de ne pas être traité.

Pour résoudre ce problème, un mécanisme de suivi des dates déjà traitées, a été mis en place. Lors de la reprise d'extraction, le programme récupère toutes les dates déjà traitées, identifie les périodes manquantes, puis relance l'extraction pour chacune de ces périodes. Néanmoins, cette approche présentait des limitations, car pour une date donnée, plusieurs produits pouvaient coexister, rendant complexe la distinction entre les données déjà traitées et les nouveaux produits nécessitant une nouvelle extraction.

Pour remédier à cette limitation, une approche alternative a été adoptée. Plutôt que de regrouper l'ensemble des données dans un unique fichier CSV, le programme génère dorénavant un fichier CSV distinct pour chaque produit. Les noms de ces fichiers CSV sont dérivés de la convention de nommage des produits, comportant l'ajout du nom du shapefile précédant le nom du produit, suivi de l'indice et du masque nuage correspondant.

Cette approche présente divers avantages. Cette modification vise à faciliter l'identification des produits déjà traités. En effet, en utilisant le nom complet conforme à la convention de nommage, elle simplifie le repérage des produits déjà traités, éliminant la nécessité d'ouvrir chaque fichier CSV, grâce à une correspondance directe entre le nom du fichier CSV et le produit extrait.

Un script de compilation des données a été élaboré pour effectuer une comparaison entre les lignes des fichiers CSV actuels et les lignes de données du fichier CSV initial. Cette étape a été suivie d'une phase de test visant à détecter d'éventuels problèmes de traitement ou d'erreur. La vérification initiale, réalisée ligne par ligne, s'est avérée trop chronophage. Pour optimiser ce processus, des DataFrames et des opérations vectorielles ont été utilisés pour extraire exclusivement les lignes présentant des divergences. Au cours de cette phase, nous avons identifié le problème de gestion des multiples produits correspondant à la même date, mentionné précédemment, ainsi que quelques anomalies lors de la compilation. Ces points seront traités dans la section suivante. Toutefois, aucun problème n'a été détecté au niveau des données extraites, ce qui m'a conforté dans la méthode employée.

Par ailleurs, un paramètre a été intégré permettant de forcer le programme à ne pas prendre en compte les données déjà présentes. Cette fonctionnalité donne à l'utilisateur la possibilité de contourner la reprise intelligente pour forcer si besoin un retraitement des données, offrant ainsi un contrôle supplémentaire sur le processus. Cette mise en place pourrait s'avérer particulièrement utile pour adapter le programme à des situations spécifiques ou pour exclure des données potentiellement indésirables, améliorant ainsi la robustesse et la personnalisation de l'outil d'extraction des données.

Gestion du "no data"

Enfin, il est important de souligner que certains produits images, bien que contenant des zones d'intérêt vectorielles, ne contiennent pas toujours de données exploitables, notamment en raison de la présence de nuages dans ces zones. C'est ce qu'on appelle le "no data". Dans le cas où aucune zone d'intérêt d'un produit n'est exploitable, aucun fichier CSV n'est généré puisqu'il serait vide, le nom du produit n'est donc pas considéré comme ayant été traité. Afin

d'éviter de retraiter ces données "no data" à l'avenir, un fichier CSV spécifique, répertoriant les noms des produits contenant du "no data", est créé. Cette approche s'avère particulièrement cruciale lorsque certains shapefiles présentent un pourcentage élevé de "no data", atteignant parfois 40 à 50 %, ce qui représente un temps de traitement considérable. Le fichier CSV "no data" offre ainsi une utilité essentielle pour la gestion des données non traitées. La détection du "no data" s'effectue en repérant les produits présents dans la liste des produits répondant aux critères de traitement, mais qui ne sont pas finalement présents dans le dossier "raw data" à la fin du processus d'extraction.

Organisation du dossier "raw_data"

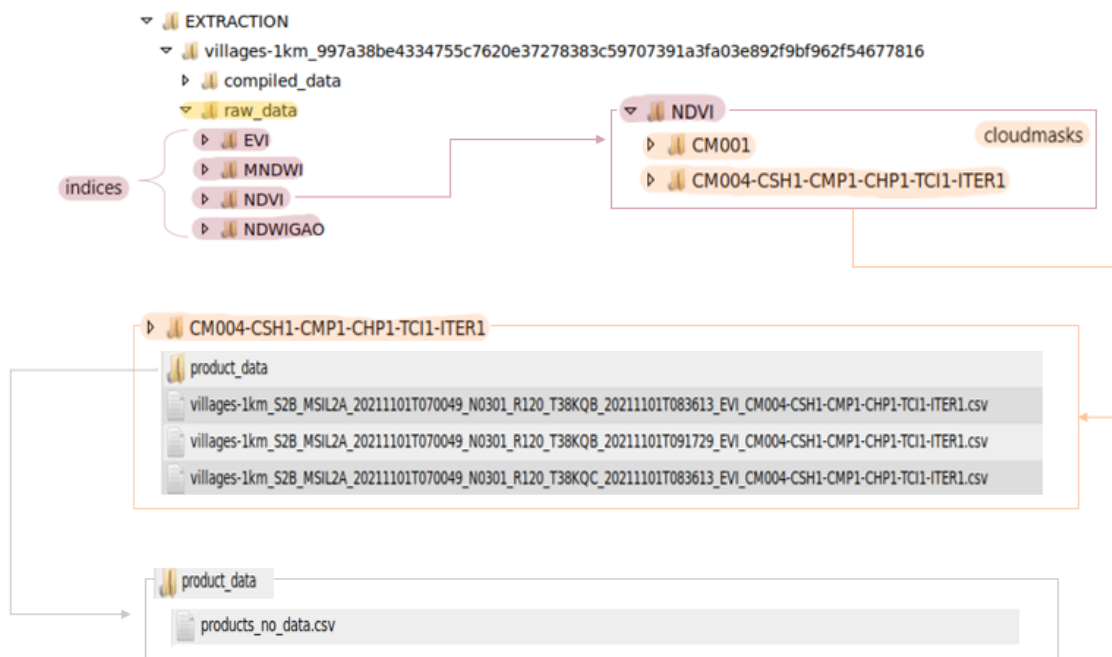


FIGURE 3.4 – Vue schématique du dossier raw_data

Pour assurer une organisation claire au sein du sous-dossier "raw_data" de l'arborescence d'extraction, chaque indice pour lequel des données ont été extraites dispose de son propre sous-dossier. À l'intérieur de ces sous-dossiers, on retrouve des dossiers distincts correspondant aux masques nuages utilisés, suivis des données organisées par produit. Cette structure hiérarchique, schématiser sur la figure 3.4, facilite la localisation et la gestion des données extraites.

3.2.6 Données compilées

Les données compilées sont les données finales exploitées par les utilisateurs. Ces données résultent de la fusion des différents fichiers CSV distinct pour chaque produit.

Il a été décidé de ne pas systématiquement effectuer une compilation complète des données, mais plutôt de permettre à l'utilisateur de choisir entre une compilation des données journalière, hebdomadaire, mensuelle, personnalisée entre deux dates précises, ou une compilation totale.

Le code d'extraction ayant subi des modifications, et afin de faciliter la transition vers le nouveau code sans perturber les données déjà exploitées, il était essentiel de garantir la concordance avec les données du CSV initial. Suite à la vérification des données grâce au script

réalisé, des différences ont été identifiées, notamment au niveau de la gestion des doublons. Une analyse approfondie a permis de conduire à l'amélioration des critères de gestion des doublons et d'assurer une compilation plus précise et conforme aux besoins du projet.

Gestion des doublons

Il est important de comprendre que certains polygones peuvent être présents sur plusieurs produits pour diverses raisons.

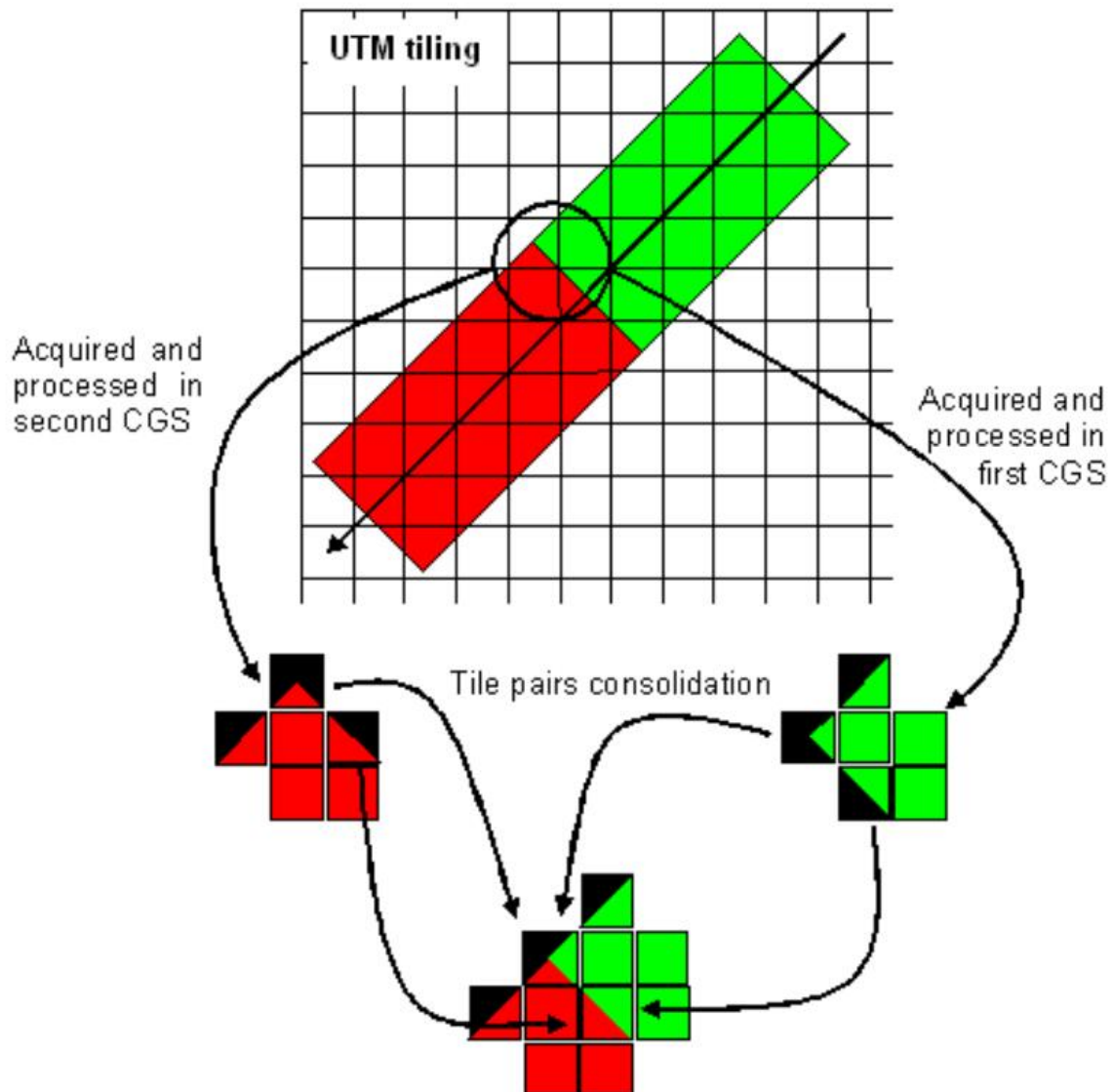


FIGURE 3.5 – Coupure des produits SENTINEL-2 [Crédits :[3]]

- **Découpage N/S** : Les acquisitions de données sont généralement constituées de bandes de données provenant de plusieurs stations de réception au sol, ce qui peut entraîner une coupure Nord/Sud au niveau des images. En d'autres termes, la bande de données ne s'arrête pas directement au début ou à la fin d'une tuile, générant ainsi deux produits distincts pour la même tuile et la même date. De plus, il est important de rappeler que la capture de données par les satellites se fait le long d'une trajectoire orientée Sud vers le Nord, c'est pourquoi cela s'appelle une coupure N/S. La figure figure 3.5 illustre la coupure dans l'acquisitions des données. On peut observer des tuiles composées de

zones en rouge et en vert, ce sont des données qui viennent de deux stations de réception différentes. Cela indique la présence de deux produits différents mais acquis le même jour et au même moment pour une même tuile. Ces deux produits se chevauchent, ce qui signifie qu'un polygone peut être présent sur les deux produits différents si son emplacement se situe dans la zone de chevauchement.

- **Chevauchement** : Nous avons mentionné dans la partie précédente qu'il y avait un chevauchement de 5km entre chaque tuile. Si un polygone se trouve dans la partie d'une tuile qui chevauche une autre, des informations sur ce polygone seront disponibles sur les deux tuiles.
- **Versions différentes** : Plusieurs versions de produits peuvent coexister en raison de l'évolution de la base de traitement des images, également appelée "processing baseline", tout au long de la mission SENTINEL-2. Les processing baselines sont des fichiers de configuration qui subissent des mises à jour au fil du temps. Ainsi, les produits qui ont déjà été traités peuvent être soumis à un nouveau traitement à mesure qu'une nouvelle processing baseline devient disponible.

La présence de plusieurs produits associés à un même polygone se traduit par la génération de multiples lignes de données pour ce polygone à une date donnée dans le fichier CSV compilé, entraînant ainsi des doublons d'information. Afin de remédier à cette situation, une étape de tri des doublons a été intégrée lors de la compilation des données. Cette procédure vise à maintenir uniquement la ligne la plus pertinente en se basant sur divers critères. Les critères initiaux de tri dans le CSV impliquaient l'ordonnancement des lignes selon une séquence déterminée, en suivant une priorité spécifique :

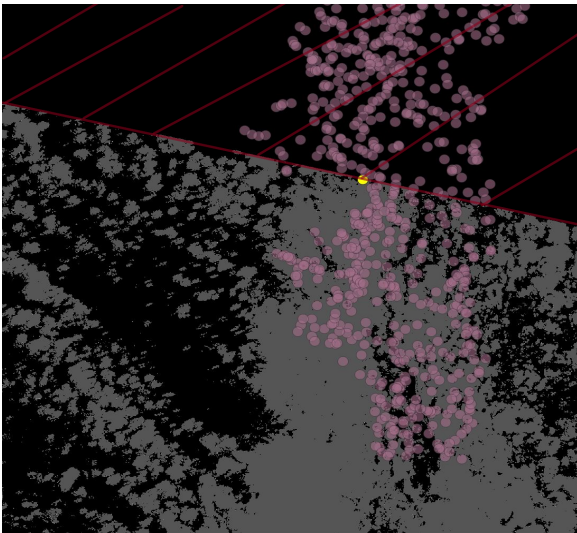
- **identifiant du polygone** : permet de limiter le tri aux polygones identiques.
- **date** : permet de limiter le tri aux polygones qui ont la même date.
- **nombre de pixels** : conservation de la ligne avec le plus de pixels de données, c'est à dire le moins de "no data" possible.
- **écart type** : conservation de la ligne avec l'écart type le plus faible.

Par la suite, les doublons sont éliminés en conservant uniquement la dernière occurrence pour chaque paire date-polygones, assurant ainsi une représentation optimale des données dans le CSV compilé.

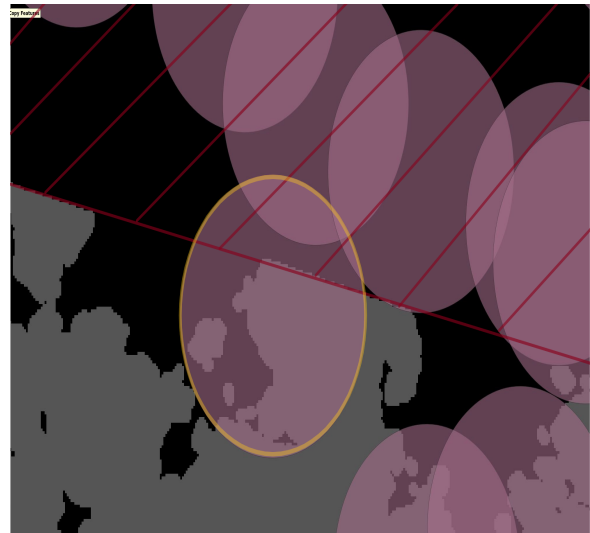
Néanmoins, ce processus de tri ne tient pas compte de la coexistence de deux produits présentant une différence au niveau de la processing baseline. Les versions actualisées des produits offrent des informations plus précises, même si la quantité de pixels contenant des données de polygones peut être réduite en raison d'une détection améliorée. Pour résoudre cette problématique, un critère de tri basé sur la processing baseline a été introduit. Ce critère prend alors le dessus sur la considération du nombre de pixels de données dans le polygone.

Le tri initial du CSV produisait des résultats différents de ceux obtenus lors de la compilation actuelle. Il a été nécessaire de comprendre la raison derrière cette divergence, qui correspondait à un cas particulier ou du moins à une situation que nous n'avions pas anticipée comme source potentielle de problème.

Ce scénario particulier se présente lorsque 3 produits images différents existent. On a donc extractions différentes ont été réalisées. Les deux premières en raison d'une coupure N/S de l'image, et la seconde un processing baseline différent sur l'image coupée au Nord. Cette situation est illustré de la figure 3.6 à la figure 3.8.



(a) Position du polygone 60

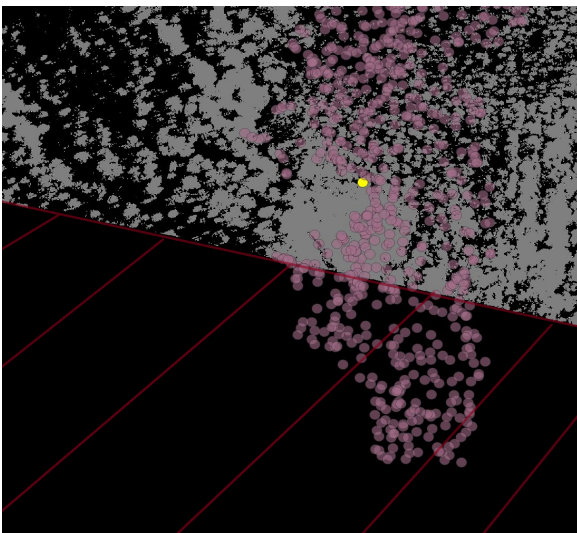


(b) Zoom sur le polygone 60

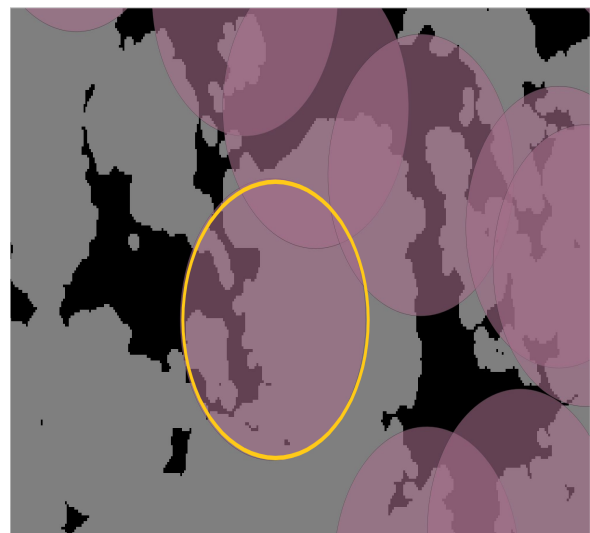
FIGURE 3.6 – S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729

Legend : Le noir représente les données, le gris représente les nuages, et le hachuré rouge représente la coupure c'est donc du "no data".

Sur la figure 3.6 ci dessus, on peut apercevoir que l'image est coupé N/S. Le polygone se trouve à la limite de la coupe et n'est donc pas complet.



(a) Position du polygone 60



(b) Zoom sur le polygone 60

FIGURE 3.7 – S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T083613

Legend : Le noir représente les données, le gris représente les nuages, et le hachuré rouge représente la coupure c'est donc du "no data".

Sur la figure 3.7 ci dessus, on peut apercevoir que l'image est coupé N/S. C'est la deuxième partie de la figure 3.6 Grâce au chevauchement de la coupe, le polygone est complet.

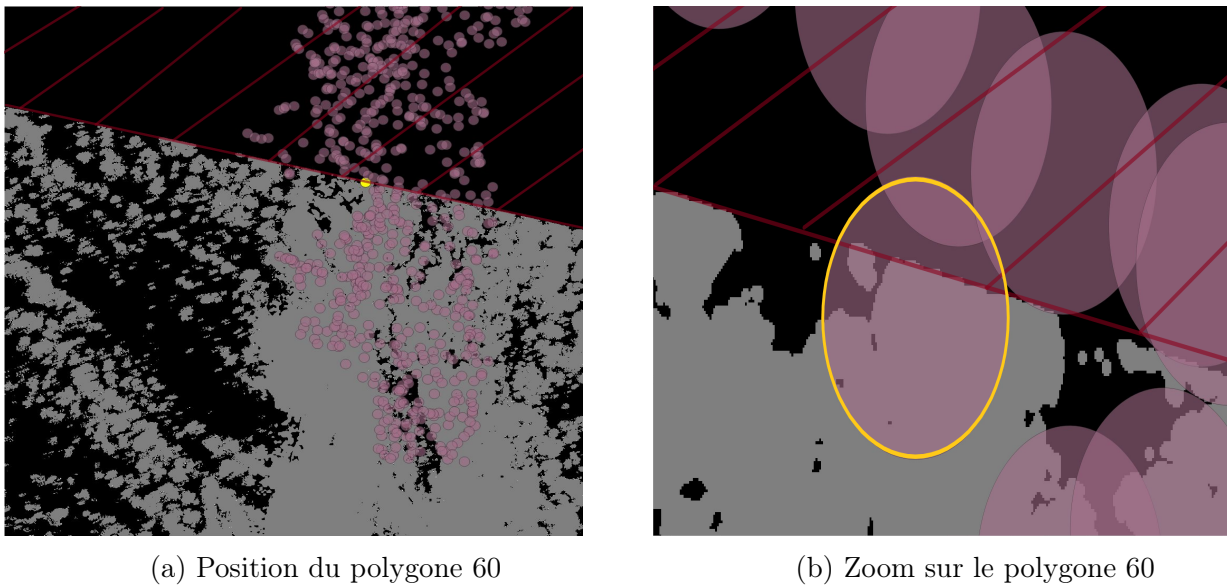


FIGURE 3.8 – S2B_MSIL2A_20211101T070049_N0500_R120_T38KQB_20221230T104543

Legend : Le noir représente les données, le gris représente les nuages, et le hachuré rouge représente la coupure c'est donc du "no data".

Sur la figure 3.8 ci dessus, on peut apercevoir que l'image est coupé N/S. C'est l'image correspondant à une version supérieure (N500) de l'image de la figure 3.6. Le polygone se trouve à la limite de la coupe et n'est donc pas complet.

Il a été nécessaire d'introduire un nouveau critère de tri : la complétude d'un polygone, favorisant les polygones complets. Cette décision a été prise pour éviter le problème précédent et améliorer la pertinence de la gestion des doublons. La complétude est calculée comme suit :

$$\text{Complétude} = 1 - \frac{\text{Nombre de pixels no data} - \text{Nombre de pixels de nuages}}{\text{Nombre de pixels de données} + \text{Nombre de pixels no data}}$$

La complétude mesure la proportion de pixels valides par rapport à la somme des pixels valides et des pixels sans données. Ainsi, elle offre une indication de la couverture effective du polygone par des données exploitables, favorisant ainsi les polygones complets dans le processus de tri.

On définit un seuil à 90%, au dessous de ce seuil, on considère que le polygone n'est pas complet.

On peut voir sur les figures précédente 3 tuiles, deux sont coupés N/S en version N0301 et une autre en version N0500. Sans la données de complétudude, la version N0500 aurait été conservé. Mais maintenant, le code peut détecter que le polygone présent sur l'image N0500 est coupé tandis que celui présent sur l'image N0301 de la figure figure 3.7 n'est pas coupé. On a donc une information plus complète sur ce produit, c'est donc celui ci qui va être conservé.

Les critères finaux de tri des doublons dans le CSV impliquaient l'ordonnancement des lignes selon une séquence déterminée, en suivant une priorité spécifique :

- **identifiant du polygone** : permet de limiter le tri aux polygones identiques.
- **date** : permet de limiter le tri aux polygones qui ont la même date.
- **compeltude** : les polygones complets sont prioritisés
- **processing baseline** : version la plus récente udu produit
- **nombre de pixels** : conservation de la ligne avec le plus de pixels de données, c'est à dire le moins de "no data" possible.
- **écart type** : conservation de la ligne avec l'écart type le plus faible.

Problème lors de la recopie des données

Par la suite, des différences à la vérification des données entre le CSV initial et les données actuelles, mais elles étaient liées à des problèmes d'arrondis lors de la manipulation des données dans les DataFrames. Ces problèmes ont été résolus en forçant les valeurs à être des chaînes de caractères. Cependant, des tout en conservant les données numériques pour les calculs de complétude.

Mise à jour intelligente des données compilées

Un code a été élaboré pour améliorer l'efficacité du processus de compilation des données. Plutôt que de relire tous les CSV de tous les produits pour les recompiler systématiquement, le code identifie d'abord les produits qui ont déjà été compilés. Il procède ensuite à la complétion sélective des fichiers CSV des données déjà compilées. Cependant, cette fonctionnalité a été retirée ultérieurement pour permettre la mise en œuvre du processus de lissage des données, que nous expliquerons dans la section suivante.

Organisation du dossier "compiled_data"

Pour assurer une organisation claire au sein du sous-dossier "compiled_data" de l'arborescence d'extraction, la structure du dossier ressemble à celui du dossier "raw_data". Chaque indice pour lequel des données ont été extraites dispose de son propre sous-dossier. À l'intérieur de ces sous-dossiers, on retrouve des dossiers distincts correspondant aux masques nuages utilisés, suivis des données organisées par produit. Cette structure hiérarchique facilite la localisation et la gestion des données extraites.

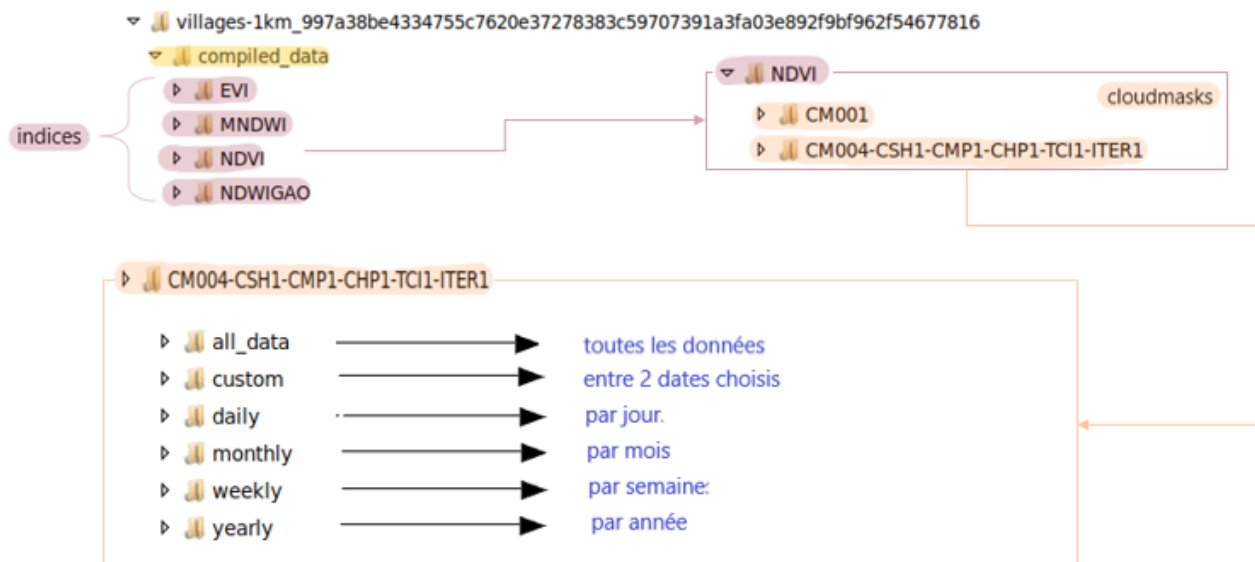


FIGURE 3.9 – Vue schématique du dossier compiled_data

3.2.7 Présentation du travail effectué

Dans le cadre des améliorations apportées, j'ai réalisé une présentation visant à partager l'état d'avancement et les détails de mon travail avec les autres membres de l'équipe de déve-

loppement. Cette étape revêtait une importance particulière pour assurer une communication efficace au sein de l'équipe.

L'objectif principal de cette présentation était de présenter les différentes phases de mon travail, les choix de conception effectués, ainsi que les solutions techniques mises en place. Cela a offert à mes collègues une compréhension approfondie des décisions prises et du fonctionnement du code que j'avais développé. L'initiative visait à faciliter l'intégration harmonieuse du code dans l'ensemble du projet en rendant son utilisation et ses implications claires et accessibles à l'équipe.

La présentation a également constitué une opportunité de m'exercer à la communication orale, en présentant de manière claire et concise les aspects techniques de mon travail. Les retours obtenus lors de cette séance ont été précieux, me permettant d'ajuster certains points et d'améliorer la qualité de mon travail.

3.2.8 Intégration du lissage temporel

L'ajout d'un lissage temporel visant à réduire le bruit présent dans les données est une étape importante dans le processus de traitement des données. Cette technique d'interpolation visait à améliorer la cohérence des données disponibles sur le catalogue en ligne de SEAS OI.

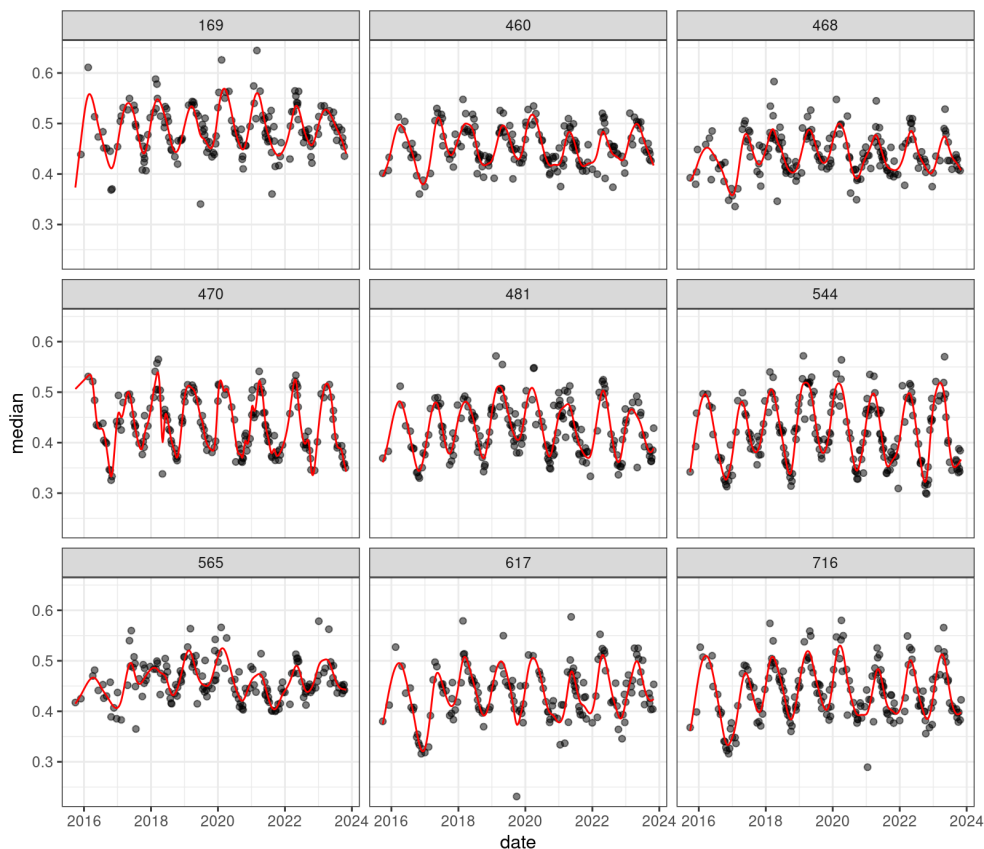


FIGURE 3.10 – Exemple d'un lissage temporel réalisés sur différents polygones d'extraction

Legend : Les données brutes sont représentées par les semis de points noirs. La courbe rouge représente le lissage.

Initialement, le lissage temporel était réalisé après les extractions et la mise à disposition des fichiers CSV, à l'aide du logiciel R. Mon objectif était de transposer ce processus en Python.

Cependant, les différences entre les fonctions disponibles dans les deux langages ont introduit des variations dans les résultats obtenus.

En raison de contraintes de temps, je n'ai pu prendre en charge cette tâche que vers la fin de mon stage, ce qui a limité le temps alloué pour sa réalisation complète. Bien que j'aie réussi à adapter le code R en Python, l'étape d'interpolation nécessaire pour obtenir les résultats souhaités était toujours en cours de développement à la fin de mon stage.

Il est important de noter que cette mission a représenté un défi, notamment en raison de mon inexpérience préalable en matière d'interpolation. Malgré cela, j'ai pu ajuster le code existant pour qu'il fonctionne dans un environnement Python. Cependant, pour parvenir à la version finale du lissage temporel, des modifications supplémentaires au niveau des fonctions mathématiques d'interpolation sont nécessaires pour obtenir les résultats attendus.

Malgré les obstacles rencontrés, cette expérience m'a permis d'approfondir ma compréhension des techniques d'interpolation temporelle. Bien que le travail ne soit pas complètement achevé, il a posé les bases nécessaires pour la poursuite du développement et de l'optimisation du processus de lissage temporel dans le contexte spécifique de SEAS OI.

3.2.9 Fichier log

Au cours de l'ensemble du processus d'extraction, des journaux (logs) sont affichés sur le terminal pour permettre le suivi en temps réel des opérations en cours. Cependant, un problème majeur se posait, car ces logs n'étaient pas enregistrés, ce qui compliquait les exécutions régulières du code en mode routine. À la demande de l'équipe de développement, j'ai donc initié la création de fichiers de log dans un répertoire dédié pour consigner ces informations.

La création de fichiers de log a été mise en œuvre pour enregistrer systématiquement les messages affichés sur le terminal au cours des différentes étapes du processus d'extraction. Cela s'est avéré essentiel pour la traçabilité des opérations et la détection rapide de problèmes éventuels.

Chaque fichier de log est associé à une exécution spécifique du code d'extraction, permettant ainsi un suivi chronologique des opérations. Ces fichiers contiennent des informations détaillées telles que les étapes effectuées, les éventuelles erreurs survenues, et d'autres données pertinentes pour faciliter le débogage et l'optimisation ultérieure du code.

Cela renforce la robustesse du système, en garantissant la disponibilité d'un historique complet des opérations effectuées lors des exécutions du code en routine.

3.2.10 Modification de la classe Produit

Une modification a été apportée à la classe Produit pour résoudre un problème lié à l'absence de données sur le pourcentage de nuages dans les fichiers CSV. Le problème découle du fait que je n'avais pas accès aux données L1C pour les extractions, ce qui contribuait à alléger la charge de travail de ma machine virtuelle. Récemment, les métadonnées sur le pourcentage de nuages ont été copiées des dossiers L1C vers les dossiers L2A, afin d'éviter de parcourir deux fois la liste des produits. Cela permet aussi d'envisager la suppression des fichiers L1C, qui prennent énormément de place.

Cependant, un nouveau défi est apparu car les fichiers n'étaient pas nommés de manière identique. Afin de résoudre cette problématique, des modifications ont été apportées à la classe Produit, permettant ainsi de lire correctement les fichiers correspondants et assurant la cohérence des données.

3.2.11 Mise en production

Pour implémenter ces ajustements, une nouvelle extraction des données existantes a été lancée. Ce processus a pris du temps, car la reprise intelligente n'était pas opérationnelle initialement, étant donné qu'aucune donnée n'avait encore été extraite. Une fois le code mis à jour, il a été déployé en production. Les mises à jour des extractions, après cette amélioration, permet d'économiser de plusieurs jours de calculs. Bien que cela ait nécessité un nouveau téléchargement complet, la rapidité des mises à jour s'est avérée être un avantage non négligeable.

De plus, la compilation des données offre aux utilisateurs la possibilité de télécharger des fichiers de données plus petits au lieu d'un lourd fichier couvrant une période temporelle complète. Auparavant, lorsqu'il y avait une mise à jour des données, les utilisateurs devaient télécharger l'intégralité du fichier CSV, même pour les données déjà présentes. Cette amélioration a des avantages concrets, notamment pour les utilisateurs, qui peuvent maintenant profiter de données compilées, évitant ainsi les contraintes liées au téléchargement de gros fichiers CSV.

3.3 Mise en place d'une routine d'extraction

Sen2Chain utilise des tâches (jobs) pour exécuter des opérations complètes de traitement sur n'importe quelle tuile.

3.3.1 La classe Job

La classe Job est une classe pour créer, éditer, sauvegarder et lancer des routines de travail. Une routine (job) permet de télécharger les L1C, de traiter les L2A, des masques nuageux et des indices radiométriques en une seule fois, pour une ou plusieurs tuiles.

Les jobs peuvent être planifiés avec cron, et une option de journalisation est disponible.

Un job est composé de :

- Un fichier de configuration
- Un fichier .py est requis si le job est programmé avec cron. Ce fichier .py exécute la commande `run()`, qui correspond à l'exécution des instructions spécifiées dans le fichier de configuration.

Le fichier de configuration du job contient les paramètres globaux du job ainsi que les tâches. Tous les paramètres des fonctions de Sen2Chain peuvent être spécifiés, permettant ainsi la production de plusieurs produits en une seule opération. Chaque tâche est associée à une seule tuile, impliquant que pour télécharger et traiter plusieurs produits de tuiles, le fichier de configuration doit comporter le même nombre de lignes de tâches. Il est recommandé d'éditer manuellement le fichier de configuration en cas de mise en place d'un job avec plusieurs tuiles, car l'édition en ligne de commande est destinée à une tâche à la fois.

Exemple du contenu d'un fichier de configuration

```
logs = True
timing = 0 20 * * *
provider = peps
tries = 2
sleep = 4
nb_proc = 18
copy_L2A_sideproducts = False
clean_before = True
clean_after = True
```

```
tile;date_min;date_max;max_clouds;l1c;l2a;cloudmasks;indices;remove;comments
40KCB;;today;80;True;True;CM004-CSH1-CMP1-CHP1-TCI1-ITER1;NDVI/NDWIGAO/MNDWI;l1c/l2a;
```

Explications

La première partie contient des paramètres globaux pour l'exécution du job :

- `logs`: Active (True) ou désactive (False) les journaux de la tâche.
- `timing`: Fréquence de la tâche lorsqu'elle est ajoutée à `cron`, en format `cron`.
- `provider`: Fournisseur pour télécharger les produits L1C, par défaut `peps`, valeurs possibles : `peps` | `scihub`.
- `tries`: Nombre d'essais que le téléchargement devrait effectuer avant de s'arrêter, pour télécharger des produits `OFFLINE`.
- `sleep`: Temps en minutes à attendre entre les essais de téléchargement.
- `nb_proc`: Nombre de cœurs CPU à utiliser pour cette tâche, par défaut 8.
- `copy_L2A_side_products`: Duplique `msk_cldprb_20m` et `scl_20m` du dossier L2A vers le dossier `Cloudmask` après la production de L2A. Utile si l'on prévoit de supprimer L2A pour économiser de l'espace disque, mais que l'on souhaite conserver ces 2 fichiers pour la génération de masque de nuages et une meilleure extraction. Valeurs possibles : `True` | `False`.
- `clean_before` / `clean_after`: Nettoie les fichiers corrompus dans le dossier de la tuile, avant et/ou après l'exécution de la tâche : `True` | `False`.

La deuxième section du fichier de configuration représente une liste de tâches qui seront traitées sur une seule tuile lors de l'exécution de la tâche :

- `tile`: Identifiant de la tuile, format `-XXX`, la ligne de commentaire utilise `!` avant le nom de la tuile.
- `date_min` / `date_max`: Date de début et de fin de la tâche, les valeurs possibles incluent une chaîne vide (2015-01-01 sera utilisé), n'importe quelle date ou `today-xx` (xx jours avant aujourd'hui).
- `max_clouds`: Couverture nuageuse maximale à considérer pour le téléchargement des images.
- `l1c` / `l2a`: Télécharge L1C : `True` | `False`, Calcule L2A avec `Sen2Chain` : `True` | `False`.
- `cloudmasks`: Masque(s) de nuages à calculer et masquer l'indice (s). Valeurs possibles : de `none` (False) à plusieurs masques de nuages : `False` | `CM001/CM002/etc`.
- `indices`: `False` | `All` | `NDVI/NDWIGAO/etc`.
- `remove`: Supprime L1C téléchargé et/ou L2A produit une fois la tâche terminée, valeurs possibles : `False` | `True`

3.3.2 Objectifs

Notre objectif est de développer une classe `JobExtraction` en parallèle à la classe existante `Job`. Cette nouvelle classe est spécifiquement conçue pour superviser les routines d'extraction. Il est important de souligner que certaines extractions sont réalisées de manière récurrente sur les mêmes shapefiles. Cette approche s'explique par le besoin constant d'avoir des données sur ces zones d'intérêt, répondant ainsi à diverses demandes ou projets.

3.3.3 Classe JobExtraction

La classe `JobExtraction` permet de gérer les extractions de routine. Tous les paramètres des fonctions de la classe `TimeSeries` de `Sen2Chain` peuvent être spécifiés.

3.3.4 Fichier de configuration

Le fichier de configuration de `JobExtraction` est une adaptation du fichier d'extraction de la classe `Job`. Cette ressemblance permet à l'utilisateur une facilité d'adaptation. Ici, chaque ligne correspond à un shapefile à traiter.

3.3.5 Utilisation du cron

Explication du cron

Le cron est un utilitaire dans les systèmes d'exploitation Unix et Unix-like permettant aux utilisateurs de planifier l'exécution automatique de tâches à des moments spécifiques. Il fonctionne en arrière-plan et exécute les tâches en fonction du temps défini dans les fichiers cron.

Le fichier cron est un fichier de configuration qui contient les détails des tâches planifiées. Chaque utilisateur peut avoir son propre fichier cron pour spécifier les tâches à exécuter selon son propre calendrier.

Le format de spécification du temps dans le fichier cron est le suivant :

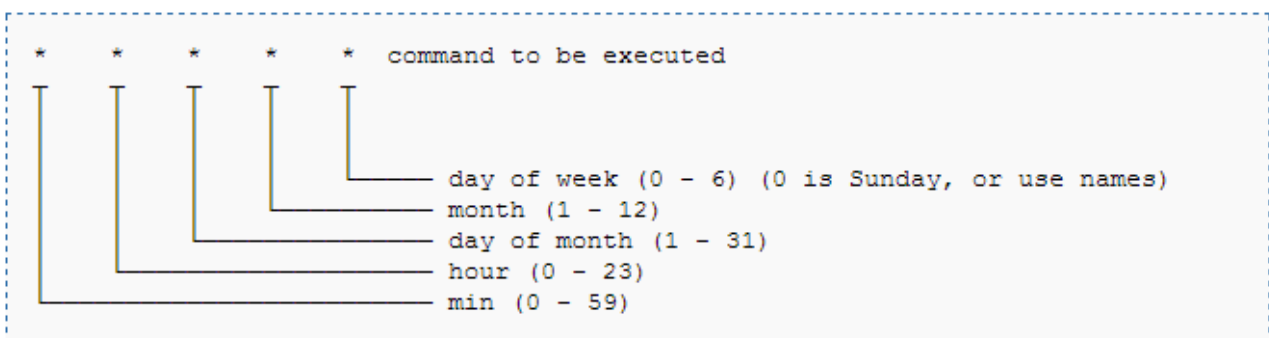


FIGURE 3.11 – Spécification dans le cron

- `minute` : Le numéro de la minute (de 0 à 59).
- `heure` : Le numéro de l'heure (de 0 à 23).
- `jour` : Le jour du mois (de 1 à 31).
- `mois` : Le mois (de 1 à 12).
- `jour_de_la_semaine` : Le jour de la semaine (de 0 à 6, où 0 est dimanche).
- `command to be executed` : La commande à exécuter.

Par exemple, une entrée dans le fichier cron suivante exécutera la commande `mon_script.sh` chaque jour à 2 heures du matin :

```
0 2 * * * mon_script.sh
```

Job dans cron

L'utilisation des fonctions `cron_enable()`, `cron_disable()`, et `cron_remove()` dans le contexte de Sen2Chain permet de gérer ces tâches planifiées dans le système cron, en activant, désactivant ou supprimant une tâche planifiée.

Il est possible d'activer un job dans cron selon la fréquence spécifiée par le paramètre `cron`, de désactiver et de supprimer un job dans cron en utilisant les fonctions `cron_enable()`, `cron_disable()` et `cron_remove()`.

3.3.6 Copie vers un serveur distant

Les résultats d'extraction sont accessibles via le catalogue en ligne de SEAS-OI, disponible sous la forme d'une machine virtuelle. Cette machine virtuelle fonctionne en mode console, dépourvue d'interface graphique. Afin de visualiser les résultats sur le catalogue en ligne, il est impératif de générer préalablement l'index correspondant.

Le catalogue en ligne de SEAS-OI se présente comme une plateforme centralisée, offrant aux utilisateurs une interface efficace pour consulter et récupérer les résultats d'extraction. Cette solution pratique facilite la gestion et la distribution des données extraites, encourageant ainsi la collaboration entre les utilisateurs de Sen2Chain et l'équipe de recherche.

Dans le fichier de configuration de `JobExtraction`, l'utilisateur peut définir s'il souhaite faire une copie vers un serveur distant. Il lui suffit donc d'entrer dans un champ l'adresse du serveur et d'avoir une clé ssh lui permettant l'identification.

3.4 Mise à jour des produits

Copernicus effectue régulièrement des mises à jour sur les images SENTINEL-2 ainsi que sur les métadonnées associées. Ces mises à jour se caractérisent par le remplacement des produits anciens par de nouveaux produits, entraînant la coexistence de deux ou plusieurs images pour la même date dans la base de données locale.

Actuellement, Sen2Chain ne dispose pas de la capacité nécessaire pour gérer plusieurs versions du même produit dans la base locale. Pour optimiser l'utilisation de l'espace disque et prévenir d'éventuelles erreurs, il est crucial de mettre en place une gestion plus efficace des images en doublons.

3.4.1 Classe Tile

La classe `Tile` est dédiée à la gestion des produits et des tuiles. Elle offre des fonctionnalités permettant d'obtenir des informations sur les tuiles, de répertorier les différents types de produits présents dans la base tel que les L1C, les L2A, les cloudmasks et les indices. De plus, la classe `Tile` permet aussi de lancer des traitements sur les produits manquants.

La classe `Tile` constitue un outil central pour explorer, surveiller et traiter les données liées aux tuiles, facilitant ainsi les opérations de gestion et de traitement des produits dans le cadre du workflow de Sen2Chain.

3.4.2 Objectifs du travail demandé

Les objectifs de cette mise à jour incluent le développement d'une fonctionnalité de détection des doublons locaux, englobant tous les produits de la base tels que L1C, L2A, cloudmasks

et indices. Cette fonction, explorera les doublons potentiels et vérifiera si ces derniers partagent la même empreinte géographique, en prenant en compte le découpage Nord/Sud. L'objectif est d'identifier précisément les vrais doublons. Un mécanisme sera ensuite mis en place pour gérer ces doublons, permettant la synchronisation des produits entre deux versions, ainsi que la suppression d'un exemplaire plus ancien ou plus récent en fonction des spécificités de la processing baseline. Il est essentiel que cette fonction de nettoyage puisse être appliquée à l'ensemble de la famille de produits, du L1C aux indices.

Les objectifs comprennent aussi la définition du comportement à adopter lors de l'arrivée d'une nouvelle version d'un produit déjà présent localement. Cette étape implique la prise de décision quant à télécharger ou non la nouvelle version, ainsi que la gestion appropriée de la version précédente. En cas de coexistence de deux versions d'un produit en ligne, les objectifs sont de déterminer la marche à suivre. Pour garantir une synchronisation efficace avec les produits en ligne, les objectifs incluent la mise en place d'un mécanisme de détection et de déclenchement des mises à jour de la base locale. Cela inclura le téléchargement des nouvelles versions, la production identique des produits, ainsi que la suppression des anciennes versions pour maintenir la base de données locale à jour.

Les objectifs incluent l'implémentation de la possibilité de lancer régulièrement l'outil de gestion des doublons et des mises à jour sur l'ensemble de la base de données. Cette routine de maintenance périodique vise à obtenir un état des lieux complet, facilitant ainsi la détection et la correction d'éventuels doublons ou de produits obsolètes. Cette approche proactive contribuera à maintenir la base de données dans un état optimal, en assurant la cohérence des données avec les mises à jour de Copernicus et en évitant les erreurs potentielles lors des traitements ultérieurs.

3.4.3 Gestion des doublons

Tout d'abord il faut repérer les doublons déjà présent dans les données.

Identification des Doublons

La première étape consiste à comparer les noms de produits. Si les noms de deux produits sont identiques mais que les processing baselines diffèrent, la version avec la processing baseline la plus élevée est conservée. Cette logique s'applique également aux produits L1C, cloudmasks, et indices.

Cependant, un problème similaire à celui rencontré dans la section 3.2.6 peut survenir. Quand ce cas se produit, il n'est pas possible, à partir des noms de fichiers, de déterminer à quel découpage (Nord ou Sud) correspond la version. Les métadonnées ne fournissent pas explicitement cette information. Pour résoudre cela, il est nécessaire de récupérer le pourcentage de "no data" (sans nuage) dans les métadonnées et de le comparer entre les différentes versions afin de déterminer à quel produit correspond le doublon. Une fois cela fait, le doublon peut être supprimé.

Pour les cloudmasks, indices et L1C, où la méthode précédente n'est pas applicable, une approche plus simple est adoptée. Si deux découpages N/S sont présents avec une version plus élevée pour l'un d'eux, sans pouvoir identifier à quel découpage cela correspond, les deux versions sont conservées.

La suppression des doublons revêt une grande importance car elle libère considérablement d'espace et améliore les performances.

Synchronisation des Différents Dossiers

Une fois les doublons identifiés, il est essentiel de les supprimer, ainsi que tous les produits associés. Par exemple, si un produit en doublon est repéré dans le dossier L2A, cela signifie que deux versions de ce produit sont disponibles et que la version obsolète doit être supprimée à tous les niveaux : L1C, L2A, dans les cloudmasks et les indices.

En outre, une synchronisation régulière des différents dossiers est cruciale pour la gestion des données. Les familles de produits sont donc régulièrement testées pour vérifier si tous les dossiers sont synchronisés et si les produits sont à jour.

Suppression des Doublons

Pendant les phases de test, il n'était pas possible de procéder directement à la suppression des produits identifiés. Par conséquent, un fichier CSV a été créé pour répertorier les doublons et leurs familles, comme illustré dans l'exemple ci-dessous (Figure 3.12).

```
S2B_MSIL2A_20211101T070049_N0500_R120_T38KQB_20221230T104543.SAFE,500,30.195576
S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T083613.SAFE,301,51.754880
S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729.SAFE,301,30.181491
"On va supprimer : ['S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729.SAFE', 301, '30.181491']"
Présent dans le dossier l1c : /DATA/S2/L1C/38KQB/S2B_MSIL1C_20211101T070049_N0301_R120_T38KQB_20211101T091729.SAFE
Présent dans le dossier l2a : /DATA/S2/L2A/38KQB/S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729.SAFE
Présent dans le dossier cloudmask : /DATA/S2/PRODUCTS/CLOUDMASKS/38KQB/S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729
Présent dans le dossier indice BIBG: /DATA/S2/PRODUCTS/INDICES/BIBG/38KQB/S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729
Présent dans le dossier indice BIGR: /DATA/S2/PRODUCTS/INDICES/BIGR/38KQB/S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729
Présent dans le dossier indice BIRNIR: /DATA/S2/PRODUCTS/INDICES/BIRNIR/38KQB/S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729
Présent dans le dossier indice IRECI: /DATA/S2/PRODUCTS/INDICES/IRECI/38KQB/S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729
Présent dans le dossier indice MNDWI: /DATA/S2/PRODUCTS/INDICES/MNDWI/38KQB/S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729
Présent dans le dossier indice NDRE: /DATA/S2/PRODUCTS/INDICES/NDRE/38KQB/S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729
Présent dans le dossier indice NDVI: /DATA/S2/PRODUCTS/INDICES/NDVI/38KQB/S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729
Présent dans le dossier indice NDWIGAO: /DATA/S2/PRODUCTS/INDICES/NDWIGAO/38KQB/S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729
Présent dans le dossier indice NDWIMCF: /DATA/S2/PRODUCTS/INDICES/NDWIMCF/38KQB/S2B_MSIL2A_20211101T070049_N0301_R120_T38KQB_20211101T091729
```

FIGURE 3.12 – Repérage de la famille de doublons

Ce fichier CSV a été utilisé pour effectuer des tests et vérifier l'identification correcte des doublons. Une fois que tous les chemins ont été récupérés, la suppression des doublons peut être initiée.

3.4.4 Classe download EODAG

La classe `S2cEodag` est un module facilitant le téléchargement d'images Sentinel-2 par le biais d'EODAG (Earth Observation Data Access Gateway). Elle s'initialise avec des paramètres tels que le nom de la tuile et le fournisseur de données. La méthode 'search' permet de rechercher des produits Sentinel-2 en ligne selon divers critères comme le type de produit, la date, et le pourcentage de couverture nuageuse.

Après la recherche, la classe offre la possibilité de filtrer les produits en fonction de critères spécifiques, comme la présence des produits L1C, L2A, des masques de nuages, ou des indices. Cela est utile afin de ne pas télécharger des produits déjà présent. Des messages de journalisation sont générés tout au long de ce processus pour informer l'utilisateur sur les actions effectuées.

3.4.5 Gestion des téléchargements

Il est nécessaire d'ajuster la gestion des téléchargements maintenant que les doublons vont être supprimés, afin d'éviter de télécharger à nouveau des produits déjà précédemment supprimés. Ceci est crucial pour optimiser le temps et éviter les téléchargements redondants. Les

produits L1C déjà présents ou ceux ayant une correspondance L2A ne doivent pas être téléchargés, compte tenu de la suppression progressive des produits L1C.

La comparaison des noms de produits L1C est effectuée, cependant, la métadonnée utilisée pour repérer les doublons L2A n'est pas disponible pour les produits L1C. Par conséquent, une vérification plus simple basée uniquement sur les versions est réalisée. Si une image est obsolète mais que ses deux découpes nord et sud ne sont pas présentes dans la version plus récente, alors le téléchargement est effectué. Cette approche vise à assurer une gestion efficace des téléchargements tout en tenant compte des contraintes spécifiques aux produits L1C.

En explorant les métadonnées, nous avons identifié des informations relatives aux angles d'incidence. Des tests sont alors réalisés pour évaluer si ces informations peuvent être utilisées pour déterminer le découpage de l'image. De plus, afin de supprimer progressivement les produits L1C et L2A une fois que les cloudmasks et indices ont été générés, les métadonnées des images à supprimer seront sauvegardées. Cela permettra une gestion méthodique de la suppression des produits une fois que toutes les étapes nécessaires auront été complétées.

3.5 Développement du rover centipede

3.5.1 Présentation du réseau Centipede RTK

Les instituts de recherche ont recours à la géolocalisation pour indexer leurs données environnementales. Cependant, la précision des systèmes de navigation par satellite, influencée par les conditions atmosphériques, ne dépasse souvent que quelques mètres. Pour pallier ce manque de précision, des systèmes de correction différentiels sont utilisés, tels que la Cinématique en Temps Réel (RTK - Real Time Kinematic), qui permet d'atteindre une précision centimétrique. Néanmoins, le RTK nécessite une base de référence à proximité du lieu de collecte, l'accessibilité à son signal, et un équipement coûteux pour son traitement.

Au cours des dernières années, des développeurs ont activement exploité des logiciels et des composants open-source/open-hardware pour concevoir des solutions géolocalisées abordables, fiables, légères et faciles à utiliser.

La Cinématique en Temps Réel (RTK) repose sur l'utilisation de mesures de la phase des ondes porteuses des signaux émis par les systèmes GPS, GLONASS ou Galileo. Une station de référence fournit des corrections en temps réel, permettant d'atteindre une précision de l'ordre du centimètre.

Cette technologie trouve des applications variées, notamment dans les relevés naturalistes de la flore et de la faune, les relevés photographiques aériens avec des drones, et l'automatisation de la conduite des véhicules agricoles pour améliorer la précision, la qualité et la traçabilité des interventions. Le RTK est également utilisé dans les systèmes de pilotage automatique, les travaux publics, l'agriculture de précision, et d'autres domaines similaires.

Dans la pratique, les systèmes RTK utilisent une station de base (récepteur fixe) dont la position est précisément connue, ainsi que plusieurs récepteurs mobiles. La station de base compare la position calculée à partir du signal GPS avec la position réelle, puis transmet les corrections nécessaires aux récepteurs mobiles. Ainsi, les unités mobiles peuvent calculer leur position relative avec une précision de quelques millimètres, tout en conservant une position absolue aussi précise que celle de la station de base. La fiabilité et la précision opérationnelles dépendent de la densité et des capacités du réseau de stations de référence.

3.5.2 Objectif

Actuellement, La Réunion bénéficie de cinq bases du réseau Centipède, dont l'une est installée à la station SEAS-OI. Un rover préinstallé (Sparkfun Surveyor) est actuellement en service. CentipedeRTK constitue un réseau collaboratif de bases GNSS ouvertes accessible à toute personne se trouvant dans la zone de couverture, et son expansion est assurée par des instituts publics, des particuliers, des acteurs privés tels que les agriculteurs, ainsi que d'autres partenaires publics. Actuellement, un unique rover préinstallé (Sparkfun Surveyor) est en service. Idéalement, l'objectif serait de déployer plusieurs de ces récepteurs pour étendre notre capacité d'acquisition de données.



FIGURE 3.13 – Vue de la situation des bases GNSS sur La Réunion

3.5.3 Déroulement

L'ensemble de la documentation est disponible sur le site de Centipede. Voici les principales étapes du processus :

- Mise à jour et programmation des composants du module Bluetooth.
- Mise à jour et programmation des composants de la carte F9P.
- Assemblage des différents composants matériels.

Malgré quelques problèmes rencontrés lors de la configuration des composants, l'assemblage a été réalisé avec succès.

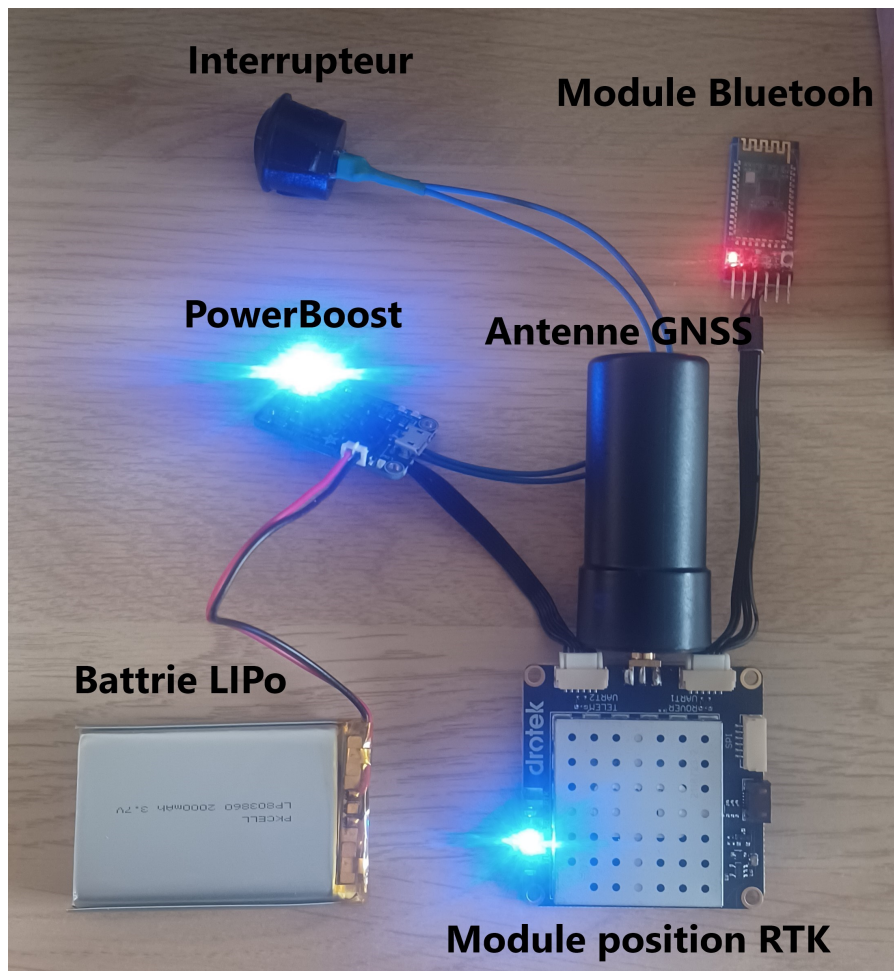


FIGURE 3.14 – Assemblage des composants du rover

3.6 Participations à d'autres activités liées à la télédétection et à l'informatique

J'ai eu l'opportunité de participer à divers événements et réunions liés à la télédétection et à l'informatique, contribuant ainsi à élargir mes horizons.

3.6.1 Rencontres Géomatiques de la Réunion (RGR - 2023)

Les Rencontres Géomatiques de la Réunion sont un rendez-vous incontournable pour les acteurs de la géomatique à La Réunion et dans la zone Océan Indien. Lors de cette édition qui s'est déroulée au domaine Moca (Saint-Denis), j'ai pu participer à un cycle de conférences passionnantes couvrant divers thèmes tels que l'aménagement du territoire, la modélisation spatiale, les risques naturels, la collecte de données géographiques, l'intelligence artificielle dans la géomatique, ainsi que le programme LiDar HD pour la cartographie en 3D de la France. Des échanges sur l'évolution des projets présentés aux éditions précédentes ont également enrichi cette expérience collaborative.

3.6.2 Restitutions du Projet G2Oi

En tant que composante essentielle du projet Grand Observatoire de l'Océan Indien, l'objectif était de mettre en valeur les données de recherche dans la région en établissant un réseau

de partenaires. Ces partenaires, présents dans les pays membres de la Commission de l'Océan Indien (COI) tels que La Réunion, les Comores, Madagascar, Maurice et les Seychelles, ont collaboré pour créer et nourrir une infrastructure de données régionale. Cette infrastructure a été conçue pour permettre aux utilisateurs finaux, qu'ils soient scientifiques, gestionnaires ou du grand public, d'accéder facilement aux données régionales via des services web disponibles sur un portail de données dédié à l'infrastructure.

Les solutions techniques adoptées ont été choisies en vue de favoriser l'interopérabilité avec d'autres initiatives similaires dans la région de l'océan Indien et au niveau mondial. À la fin du projet, une restitution ainsi que des ateliers ont été organisés pour partager les résultats et les enseignements tirés de cette initiative collaborative.

3.6.3 Assemblée générale de l'IRD

Lors de cette assemblée, l'ordre du jour comprenait des points cruciaux tels que la mise à jour sur le COMP (Le Contrat d'objectifs, de moyens et de performance) ainsi que l'opportunité pour les participants de partager quelques mots sur leurs projets. L'organisation de la représentation, les aspects administratifs, notamment la gestion des missions, ont également été abordés. Cette participation m'a offert une compréhension approfondie des activités et des perspectives de l'IRD, renforçant ainsi ma vision du paysage de la recherche et du développement dans la région.

Bilan final

Ce rapport met en avant ma contribution à l'outil Sen2Chain de SEAS-OI. Cette expérience s'est avérée non seulement enrichissante d'un point de vue technique, mais également formatrice pour le développement de compétences transversales.

3.7 Perspectives

Certaines améliorations peuvent être envisagées pour renforcer les aspects de la chaîne de traitement sur lesquels j'ai travaillé au cours du stage :

1. **Lissage sur Python** : Finaliser l'intégration du lissage en Python lors de la compilation des données. Cette étape contribuera à optimiser le processus global et à assurer une cohérence dans le traitement des informations.
2. **Gestion des doublons** : Explorer l'utilisation de l'angle d'incidence comme moyen efficace pour la gestion des doublons. Cette approche pourrait améliorer la précision du traitement en identifiant et en traitant de manière intelligente les données en double.
3. **Modifications de l'interface Sen2Extract** : Sen2Extract est une interface qui permet de réaliser les extractions temporelles. Il faudrait donc considérer des ajustements de l'interface sen2Extract au vue des changements sur la méthode d'extraction des séries temporelles.
4. **Boîte pour le rover centipede** : Réalisation d'une boîte étanche afin de permettre l'utilisation du rover.

3.8 Bilan personnel

Le stage TN09 est une première expérience professionnelle prolongée pour un étudiant UTC. C'était une aventure riche en enseignements qui a contribué de manière significative à ma croissance personnelle et professionnelle.

Le stage de 24 semaines m'a confronté à diverses situations, permettant de déceler les obstacles auxquels un ingénieur peut être confronté.

Participer à un projet en constante évolution a renforcé ma résilience et ma capacité à travailler efficacement dans des situations dynamiques, mettant en avant l'importance de l'adaptabilité dans le domaine de l'ingénierie. Les modifications fréquentes et les évolutions techniques nécessitent une approche collaborative, cela souligne la valeur de la coopération et du travail d'équipe au sein d'un projet.

D'un point de vue personnel, cette expérience a été d'un grand intérêt. La diversité des tâches et le sujet particulièrement intéressant du projet a été une source de motivation. Plonger dans ces différentes responsabilités a non seulement élargi mes compétences mais suscité un intérêt profond.

Ainsi, ce stage a été une plateforme de développement non seulement pour des compétences techniques, mais aussi pour le développement des soft skills telles que la communication, la résilience, l'adaptabilité et l'esprit d'initiative.

Cette approche a également mis en lumière les avantages de l'approche pédagogique de l'UTC, mettant l'accent sur la capacité d'apprentissage plutôt que sur la simple maîtrise des technologies. J'ai également pris conscience de l'avantage de l'approche d'enseignement à l'UTC axée sur la capacité d'apprentissage plutôt que sur la maîtrise des technologies. J'y ai appris sur mes points forts et faibles, sur mon esprit critique, et sur la nécessité de prendre des initiatives.

Cette expérience a également motivé la poursuite de ma formation et l'approfondissement de mes connaissances.

La modification d'une application existante m'a sensibilisé à l'importance des documentations techniques et des commentaires dans le code source, soulignant la nécessité d'une approche méthodique et claire dans le développement informatique.

L'expérience a élargi mes horizons en me plongeant dans la vie quotidienne professionnelle à La Réunion. Cela m'a offert une perspective plus profonde sur le monde du travail et aussi sur les enjeux spécifique du territoire réunionnais. La possibilité de réaliser ce stage dans un environnement familial a ajouté une dimension particulière à mon expérience, me permettant de redécouvrir les paysages magnifiques de l'île et profiter du cadre familiale.

Annexe

Partie 1

Le schéma figure 15 représente l'infrastructure serveur de SEAS-OI.

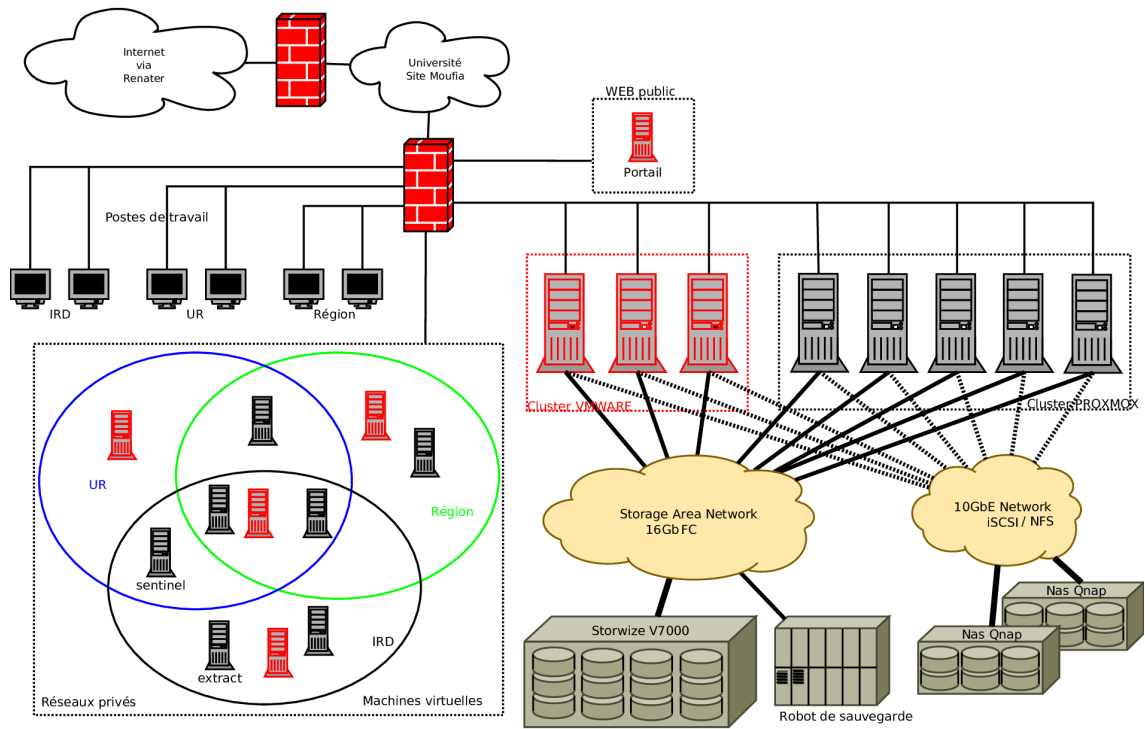


FIGURE 15 – Schema de l'infrastructure de SEAS-OI

Bibliographie

Documentation

1. WOLFF, Charlotte. *Automatisation de l'acquisition et du traitement des images Sentinel-2 pour le calcul d'indices de végétation aidant à la prévention des pics de paludisme à Madagascar*. 2017. Thèse de doct. Ecole National des Sciences Géographiques.
2. MOUQUET, Pascal ; ALEXANDRE, Cyprien ; RASOLOMAMONJY, Jaotiana ; ROSA, J ; CATRY, Thibault ; RÉVILLION, Christophe ; RAKOTONDRAOMPIANA, Solofo ; PENNOBER, Gwenaëlle. Sentinel-1 and Sentinel-2 Time Series Processing Chains for Cyclone Impact Monitoring in South West Indian Ocean. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2020, t. 43, p. 1593-1599.
3. GATTI, A ; BERTOLINI, A. Sentinel-2 products specification document. Available online (accessed February 23, 2015) <https://earth.esa.int/documents/247904/685211/Sentinel-2+ Products+ Specification+ Document>. 2013.
4. OVERLEAF. *Overleaf Learn - Online LaTeX Tutorials*. Aussi disponible à l'adresse : <https://www.overleaf.com/learn>. Consulté le 8 novembre 2023.
5. OVERLEAF. *Overleaf Learn - Inserting image*. Aussi disponible à l'adresse : https://www.overleaf.com/learn/latex/Inserting_Images. Consulté le 10 novembre 2023.
9. PEYRE, Rémi. *Utiliser BibTEX*. École normale supérieure, 2007.

Site web

6. *Sentinel-2 Operations*. Aussi disponible à l'adresse : https://www.esa.int/Enabling_Support/Operations/Sentinel-2_operations. Consulté le 24 janvier 2024.
7. « *Le Réseau Centipede RTK* ». *Centipede RTK*. Aussi disponible à l'adresse : <https://docs.centipede.fr/>. Consulté le 2 janvier 2024.
8. *Présentation de l'IRD à la Réunion | Site Web IRD*. Aussi disponible à l'adresse : <https://www.ird.fr/la-reunion/presentation>. Consulté le 1 décembre 2023.
10. « *Explore Copernicus Satellite Missions* ». *Sentinel Online*. Aussi disponible à l'adresse : <https://copernicus.eu>. Consulté le 24 janvier 2024.
11. *Understanding Sentinel-2 Satellite Data*. Aussi disponible à l'adresse : <https://eox.at/2015/12/understanding-sentinel-2-satellite-data/>. Consulté le 03 décembre 2023.
12. « *La Réunion/Mayotte* ». *UMR Espace-Dev*. Aussi disponible à l'adresse : <https://www.espace-dev.fr/portfolio-items/reunion/>. Consulté le 2 décembre 2023.