

Université de Poitiers

Année Universitaire 2023-2024

U.F.R. Sciences Fondamentales et Appliquées

Master Traitement du Signal et des Images – parcours Objets
Connectés

RAPPORT DE STAGE

En vue de l'obtention du grade de Master

Détection des Inondations par des Données Satellites et Drone

Stage effectué entre le 01/04/2024 et le 30/09/2024

Réalisé par :

Nouhaila Bougattaya

Tuteur de stage :

Arnaud Revel

Enseignant référent :

David Helbert

Nom et adresse de la structure d'accueil :

Entreprise SERLI

Arnaud Revel



Sommaire

1 Présentation du projet	6
1.1 Cadre du Projet	6
1.2 Problématique	7
1.3 Etat d'art	7
1.4 Contexte du projet	9
1.5 Analyse SWOT	10
1.6 Présentation de l'existant	11
1.7 Tâches à effectuer	12
2 Méthodologie de travail	13
2.1 Organisation du Projet	13
2.2 Planning et Échéancier	15
2.3 Gestion des Risques	17
2.4 Gestion des Ressources	18
2.5 Communication et Suivi	18
3 Travail réalisé	19
3.1 Collection de données	19
3.2 Traitement d'images : seuillage + contours	20
3.3 Génération d'un fichier GeoJson	21
3.4 Récupération des données OSM	22
3.5 constitution d'une vérité terrain	23
3.6 Etat de l'art des mesures de correspondances de données géolocalisées	23
3.7 Validation de la détection	25
3.8 Amélioration des algorithmes de détection	25
3.9 Amélioration de la Détection des Zones d'Eau par Clustering (Algorithme K-Means) . .	28
3.10 Amélioration de la Détection des Zones d'Eau par Réseau de Neurones	29
3.11 Exploration de données du Drone	32
3.12 Limites de l'Étude	35

Table des figures

1	Détection de zone d'eau	9
2	Analyse SWOT	10
3	Représentation des zones d'eau aux Sablons avant et après l'inondation, sur une semaine	11
4	Extraction des contours des zones inondées aux Sablons	12
5	kanban	13
6	Organigramme de l'équipe	14
7	Phases du projet	15
8	diagramme de Gantt global	16
9	Outils de communication	19
10	Représentation dans la bande VH	21
11	Représentation dans la bande VV	21
12	L'extraction des contours des zones inondées aux Sablons	22
13	Vrai GeoJSON représentant la zone d'eau aux Sablons	22
14	Vrai GeoJSON dans Lac du Bourget	23
15	GeoJSON détectée dans Lac du Bourget	23
16	Exemple de matrice intra-class	25
17	problem de detection	25
18	Représentation dans la bande VH	26
19	Représentation dans la bande VV	26
20	Matrice de intra-class de lac Annecy	26
21	Résultat obtenu des distances entre le plus grand polygone (réalité vs détection) du lac d'Annecy.	27
22	Variation du seuil par rapport à la qualité de détection	27
23	Représentation réelle du lac Bourget	28
24	Représentation du lac (par algorithme k-means) basée sur les bandes VH et VV	28
25	Résultat final de l'algorithme k-means comparé avec le vrai lac Bourget	29
26	Données effectives du lac d'Annecy	30
27	Données labellisées du lac d'Annecy	30
28	Évaluation ROC du modèle	31
29	Main caption for all figures	32
30	transformation HSV	32
31	Application du seuillage classique sur les données Drone	33

32	Application de K-Means sur les données Drone	34
33	Problème de décalage de coordonnées	35
34	sentinelhub Données	35

Introduction générale

Mon projet de stage se concentre sur un enjeu crucial pour le secteur de l'assurance habitation : la détection des zones inondées à travers l'exploitation des données satellitaires, notamment celles fournies par les satellites Sentinel-1. Dans un contexte où les inondations représentent l'une des principales causes de sinistres pour les assureurs, il est impératif de développer des outils permettant d'identifier rapidement et avec précision les zones à risque. Ce projet ambitieux vise à transformer la manière dont les compagnies d'assurance évaluent les menaces liées aux inondations et à améliorer la gestion des réclamations.

La première étape de cette initiative consiste à optimiser l'utilisation des images satellitaires pour améliorer la fiabilité de la détection des inondations. En intégrant des techniques de pointe, telles que l'apprentissage profond, nous avons l'opportunité d'automatiser le processus de reconnaissance des zones inondées, rendant ainsi les systèmes de surveillance plus intelligents et réactifs. Cette innovation pourrait non seulement réduire le temps nécessaire pour détecter les inondations, mais aussi fournir aux assureurs des données précieuses pour évaluer les risques et adapter leurs offres en conséquence.

L'objectif ultime est de créer des solutions robustes pour la surveillance en temps réel et la gestion proactive des inondations, contribuant à une réponse plus rapide et plus efficace lors de ces événements. Ce projet aborde également les défis complexes liés à l'exploitation des données géospatiales dans le cadre des sinistres, cherchant à développer des solutions innovantes qui facilitent la prise de décision des assureurs.

Ce stage offre une occasion unique d'acquérir une expertise pointue dans l'analyse et le traitement des données satellitaires, tout en développant des compétences en innovation technologique et en gestion des risques. En allant au-delà de l'amélioration des technologies de détection des inondations, ce projet vise à apporter une contribution significative à la sécurité des habitations et à la résilience des communautés face aux catastrophes naturelles. En fin de compte, l'ambition est d'intégrer ces nouvelles méthodes dans les pratiques des compagnies d'assurance pour mieux protéger les biens et les personnes contre les impacts des inondations.

1 Présentation du projet

1.1 Cadre du Projet

Notre projet s'inscrit dans le cadre de la gestion proactive des risques liés aux catastrophes naturelles, avec un focus particulier sur les inondations, un enjeu majeur pour les assureurs et les collectivités locales. Ce projet, réalisé en partenariat avec l'entreprise Serli, spécialisée en technologies de l'information, et le client MAIF, vise à renforcer les capacités d'assurance en matière de détection et de réponse aux inondations grâce à l'exploitation des technologies satellitaires avancées, notamment les données fournies par les satellites Sentinel-1. L'objectif est de développer des outils permettant d'identifier rapidement les zones inondées, afin de réduire les impacts sur les personnes, les biens et les infrastructures, tout en optimisant la gestion des sinistres. Ce projet s'inscrit également dans les objectifs stratégiques de la MAIF en matière d'innovation technologique et de responsabilité sociale, en contribuant activement à la protection des assurés et des communautés contre les risques climatiques.

1.1.1 Membres



La MAIF (Mutuelle Assurance des Instituteurs de France) est une société d'assurance mutuelle française créée en 1934. Son siège social est situé à Niort, dans les Deux-Sèvres. Elle propose une gamme d'assurances pour les particuliers, les professionnels et les associations. D'après ses offres, il y a des solutions pour l'auto, l'habitation, la moto, la vie, la santé et l'épargne.

Serli est une entreprise française spécialisée dans les technologies de l'information, fondée en 1981 et basée à Poitiers. Elle offre une large gamme de services, notamment dans les domaines de l'innovation, du développement logiciel, de la visualisation de données, et de la formation. Forte de plusieurs décennies d'expérience, Serli collabore avec des entreprises de renom telles que Red Hat, Oracle et Google, en fournit des solutions technologiques sur mesure adaptées aux besoins spécifiques de ses clients.



1.1.2 Encadrement

Notre projet bénéficie d'un double encadrement, alliant expertise professionnelle et académique. **Arnaud Revel**, ingénieur d'études chez **SERLI**, assure l'encadrement technique au sein de l'entreprise. Avec une solide expérience en développement logiciel et en intelligence artificielle, Arnaud joue un rôle crucial dans l'orientation technique du projet, notamment en ce qui concerne l'intégration des solutions d'IA pour la détection des zones inondées. Sa connaissance approfondie des technologies de pointe et son engagement au sein de SERLI garantissent que le projet répond aux standards les plus élevés de l'industrie.

Parallèlement, le volet académique est supervisé par **David Helbert**, professeur à l'**Université de Poitiers** et expert en traitement du signal et de l'image. David apporte un soutien essentiel en matière de méthodologie de recherche et d'analyse des données. Son encadrement est particulièrement précieux pour assurer la rigueur scientifique du projet et aligner les objectifs sur les tendances actuelles en recherche appliquée. Ses nombreuses responsabilités au sein de l'université et son expérience en encadrement de projets de recherche constituent un atout majeur pour la réussite de notre projet.

Cette complémentarité entre l'encadrement industriel et académique permet de mener notre projet à bien, en alliant innovation technologique et excellence scientifique.

1.2 Problématique

Les inondations représentent des événements naturels où des étendues de terre, habituellement sèches, se retrouvent submergées par l'eau. Ces phénomènes se produisent lorsque la quantité d'eau qui s'écoule sur le sol dépasse sa capacité d'absorption ou d'évacuation. Différents facteurs, tels que des précipitations excessives, la fonte des neiges, le débordement des cours d'eau et même les submersions marines, peuvent en être à l'origine ([de la Transition écologique et de la Cohésion des territoires \[2023\]](#)).

En tant que l'une des catastrophes naturelles les plus dévastatrices, les inondations ont un impact significatif sur les populations, les infrastructures et l'environnement. Les conséquences des inondations peuvent être désastreuses, entraînant des pertes humaines, des dommages matériels considérables, et des perturbations économiques majeures. De plus, ces événements peuvent poser des problèmes de santé publique en contaminant les sources d'eau potable et en favorisant la propagation de maladies ([français de la biodiversité \[OFB\]](#)).

Dans ce contexte, la capacité à identifier rapidement et précisément les zones inondées est essentielle pour une gestion efficace des crises et une réponse adaptée. La mise en œuvre de technologies modernes, telles que les systèmes d'information géographique (SIG) et l'analyse d'images satellites, joue un rôle crucial dans la détection et la surveillance des inondations. Par exemple, ces outils permettent de cartographier les zones touchées, d'évaluer l'ampleur des dégâts, et de planifier les interventions d'urgence ([de ressources pour l'adaptation au changement climatique \[2023\]](#)).

1.3 Etat d'art

Selon la littérature, les méthodes d'extraction automatique de l'information des plans d'eau à partir d'images de télédétection [Li et al. \[2022\]](#) peuvent être principalement divisées en deux catégories : la segmentation par seuil et la classification d'image. La segmentation par seuil repose essentiellement sur les caractéristiques spectrales de l'eau, c'est-à-dire la manière dont l'eau réfléchit ou absorbe l'énergie électromagnétique, comme les ondes radar (par exemple, les bandes VV et VH). Cette méthode utilise ces propriétés pour construire divers modèles de classification et indices, comme l'indice de l'eau normalisé (NDWI), afin d'extraire les surfaces d'eau.

En revanche, la classification d'image intègre non seulement les informations spectrales, mais aussi les caractéristiques de texture et spatiales de l'image. Cette méthode exploite pleinement les informations de forme, de structure, et de texture des éléments au sol pour extraire les informations des plans d'eau. Cette approche est particulièrement utile dans le contexte d'images de basse et moyenne résolution, où les petits plans d'eau apparaissent souvent sous forme de pixels mixtes. Un pixel mixte est un pixel qui représente un mélange de plusieurs types de surfaces (comme l'eau, le sol, et la végétation), ce qui rend difficile une détection précise des surfaces d'eau uniquement par la méthode de seuil.

Avec l'avènement des satellites à haute résolution spatiale, les images contiennent des informations spatiales et texturales plus riches. Des techniques telles que la détection de texture par ondelettes et les matrices de cooccurrence de niveaux de gris (GLCM) sont souvent utilisées pour extraire les caractéristiques texturales des images. Les chercheurs ont élaboré des méthodes de classification et d'extraction en combinant des caractéristiques comme le spectre, l'espace et la texture, permettant ainsi une extraction plus fine des informations sur l'eau. Cependant, il est important de noter que des critères de classification trop complexes peuvent réduire l'efficacité et la précision de ces méthodes. Les méthodes de classification actuelles incluent les machines à vecteurs de support (SVM), les arbres de décision et les méthodes orientées objet. Avec les progrès technologiques et l'arrivée de l'ère des big data, les techniques d'apprentissage profond (Réseaux de neurones profonds ou bien Réseaux Deep) ont été largement adoptées pour l'extraction de l'eau.

Nous présentons ci-dessous quelques algorithmes notables trouvés dans la littérature [Li et al. \[2022\]](#) :

Seuillage	Description
Seuillage par bande	Méthode qui utilise une seule bande spectrale pour extraire les plans d'eau en fixant un seuil approprié. Cette méthode se base sur les caractéristiques spectrales des plans d'eau et fixe un seuil de réflectance ou de valeur de pixel pour distinguer les pixels correspondant à de l'eau des autres types de surface.
Méthode multibande	Méthode qui utilise plusieurs bandes spectrales et/ou des indices pour extraire les plans d'eau en définissant des seuils ou des règles basées sur les valeurs de ces bandes. Cette méthode est plus complexe que le seuillage par bande car elle peut utiliser plusieurs informations spectrales simultanément, ce qui permet une meilleure distinction entre les plans d'eau et les autres types de surface.

TABLE 1 – Méthodes d'extraction des plans d'eau

Méthodes d'apprentissage	Description
Support Vector Machine	Méthode basée sur l'apprentissage supervisé qui utilise des vecteurs de support pour définir un hyperplan de séparation optimal entre différentes classes, permettant ainsi de classifier les plans d'eau à partir de caractéristiques extraites des images de télédétection. Cette méthode est efficace pour traiter des ensembles de données de grande dimension et peut être utilisée avec différents noyaux pour modéliser des relations non linéaires entre les caractéristiques.
Arbre de Décision	Méthode d'apprentissage supervisé qui utilise une structure arborescente pour prendre des décisions basées sur des caractéristiques extraites des images de télédétection. Chaque nœud de l'arbre représente une caractéristique et chaque branche une décision basée sur cette caractéristique. Les arbres de décision sont faciles à interpréter et peuvent capturer des relations complexes entre les caractéristiques, mais peuvent être sensibles au surapprentissage.
Classification orientée objet	Méthode qui prend en compte la structure spatiale des objets dans les images de télédétection en regroupant les pixels similaires en objets et en utilisant ces objets comme unités de classification. Cette approche permet de mieux tenir compte du contexte spatial des objets et peut conduire à une meilleure précision de classification par rapport aux méthodes pixel par pixel.
Deep Learning	Approche d'apprentissage automatique qui utilise des réseaux de neurones profonds pour extraire des caractéristiques complexes à partir des images de télédétection et effectuer une classification automatique des plans d'eau. Les réseaux de neurones profonds peuvent apprendre des représentations hiérarchiques des données et sont capables de capturer des relations complexes entre les caractéristiques, ce qui les rend particulièrement efficaces pour la classification d'images de haute résolution.

TABLE 2 – Méthodes d'apprentissage pour l'extraction des plans d'eau

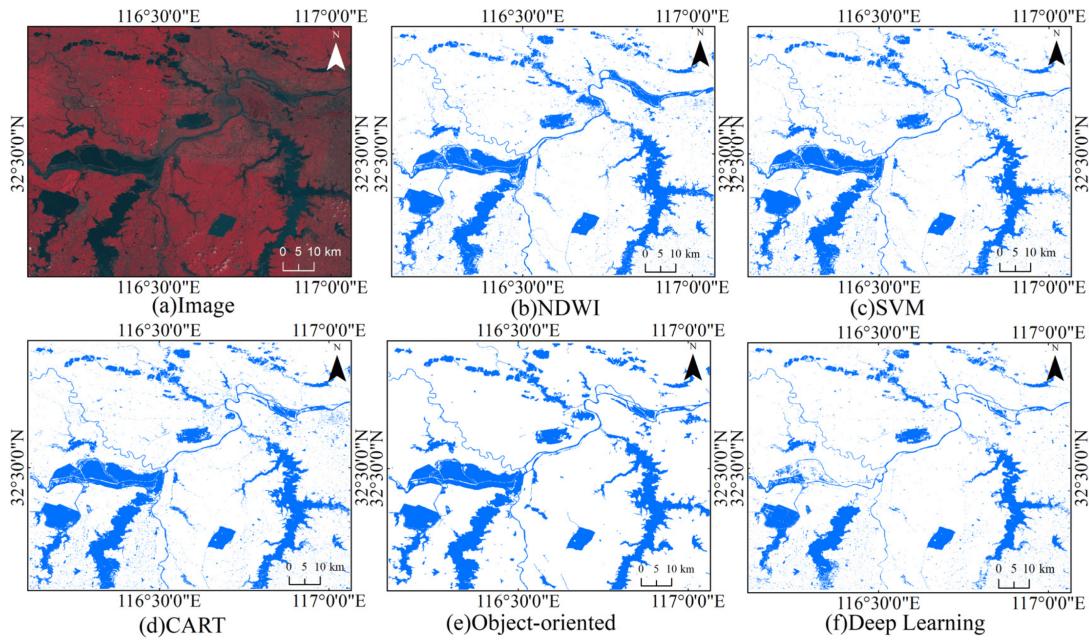


FIGURE 1 – Détection de zone d'eau

1.4 Contexte du projet

Le projet se concentre sur la reconnaissance des zones inondées à l'aide des données satellitaires, en particulier celles fournies par les satellites Sentinel-1. Initialement, des images de drones étaient envisagées, mais leur indisponibilité a conduit à se concentrer sur l'amélioration de la détection à partir des données satellitaires.

Dans cette phase initiale, l'objectif est d'optimiser l'utilisation des données de Sentinel-1. Ce projet vise à fournir une solution fiable pour la détection des inondations, contribuant ainsi à une gestion plus efficace des crises liées à ces événements.

1.5 Analyse SWOT

Voici un schéma illustrant l'analyse SWOT de notre projet, qui nous permet de mettre en évidence les forces, les faiblesses, les opportunités et les menaces auxquels nous faisons face.

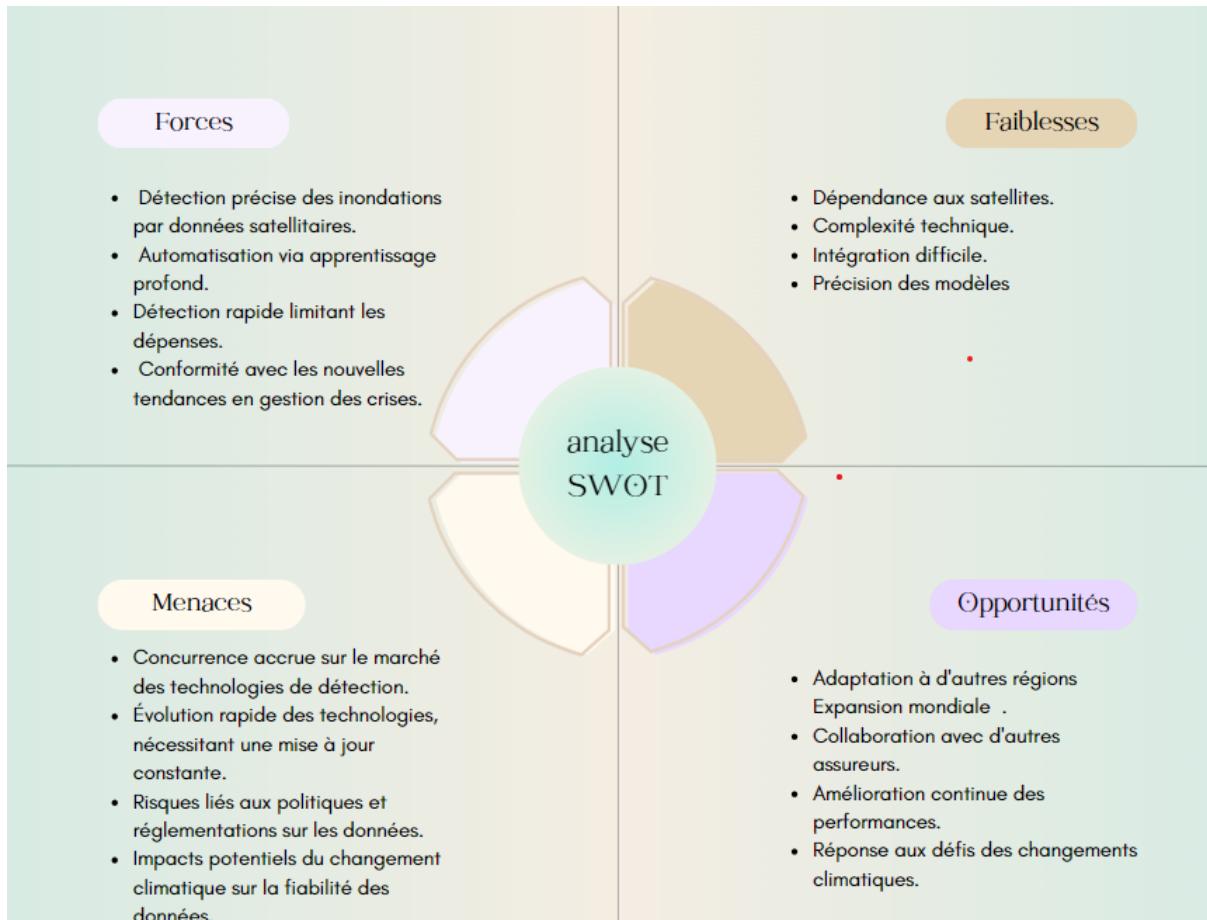


FIGURE 2 – Analyse SWOT

Pour détailler ce schéma, nous allons passer sur chaque point de l'analyse. D'abord, il y a **Forces (Strengths)** :

- Technologie de Pointe : L'utilisation avancée des données satellitaires permet une détection précise des inondations, essentielle pour une réponse rapide.
- Automatisation Intelligente : Les systèmes autonomes d'apprentissage profond améliorent l'efficacité en réduisant la nécessité d'interventions humaines.
- Réduction des Coûts : La détection rapide des inondations limite les dommages et donc les dépenses pour les assureurs.
- Alignment sur les Tendances Actuelles : Le projet adopte des innovations en gestion des catastrophes, renforçant sa position sur le marché.

Ensuite, nous avons **Faiblesses (Weaknesses)** :

- Dépendance aux Satellites : La fiabilité des données peut varier selon les conditions climatiques, ce qui peut affecter la détection.
- Complexité Technique : La mise en œuvre nécessite des compétences avancées, ce qui pourrait limiter l'équipe disponible.
- Intégration des Systèmes : L'intégration dans les infrastructures existantes peut être complexe et coûteuse.

- Risques de Précision : Les modèles d'IA peuvent générer des faux positifs ou négatifs, entraînant des erreurs de détection.

Pour les **Opportunités (Opportunities)** :

- Expansion Internationale : Adaptation possible du projet à d'autres régions vulnérables aux inondations.
- Partenariats avec les Assurances : Collaborations avec d'autres assureurs pour élargir l'impact du projet.
- Amélioration Continue : Les algorithmes peuvent évoluer pour améliorer leur précision au fil du temps.
- Réponse aux Enjeux Climatiques : Solutions adaptées aux événements climatiques extrêmes, répondant à un besoin croissant.

Enfin, pour les **Menaces (Threats)** :

- Concurrence sur le Marché : Emergence de solutions alternatives pouvant menacer notre position.
- Évolution Technologique Rapide : Nécessité de s'adapter rapidement aux nouvelles avancées technologiques.
- Risques Environnementaux : Changements climatiques pouvant affecter la fiabilité des données.
- Régulations Changeantes : Adaptation nécessaire aux évolutions des politiques gouvernementales.

Cette structure concise permet de présenter clairement chaque point de l'analyse SWOT tout en restant accessible et compréhensible, ce qui nous aide à mieux saisir les enjeux et les perspectives de notre projet.

1.6 Présentation de l'existant

Le projet s'attache à extraire les contours des zones inondées, comme l'illustre la figure 3, pour la période du 7 décembre 2023 au 14 juillet 2023. Cette figure montre la zone des Sablons en France, avec des résultats mettant en évidence les contours des zones d'eau nouvellement formées.

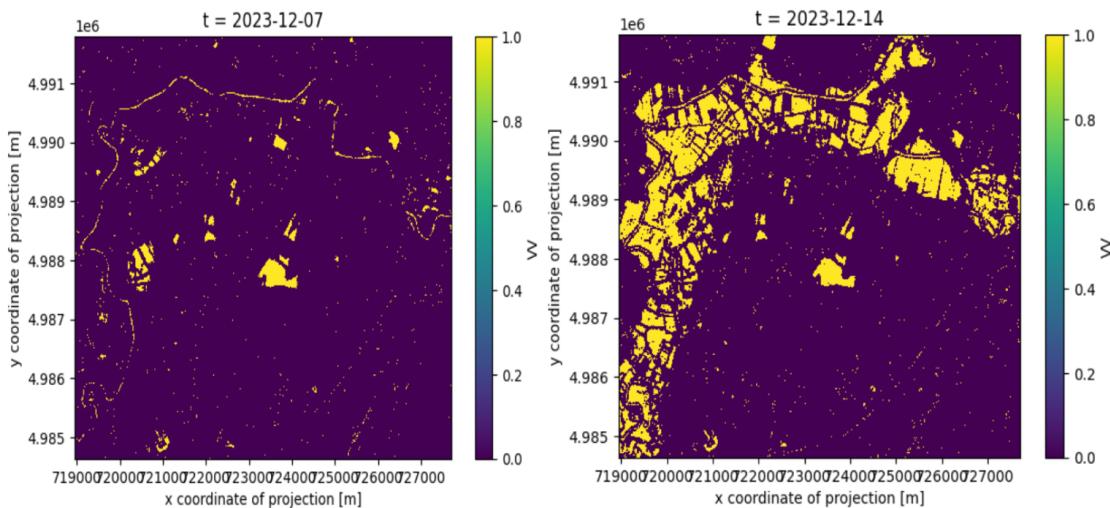


FIGURE 3 – Représentation des zones d'eau aux Sablons avant et après l'inondation, sur une semaine

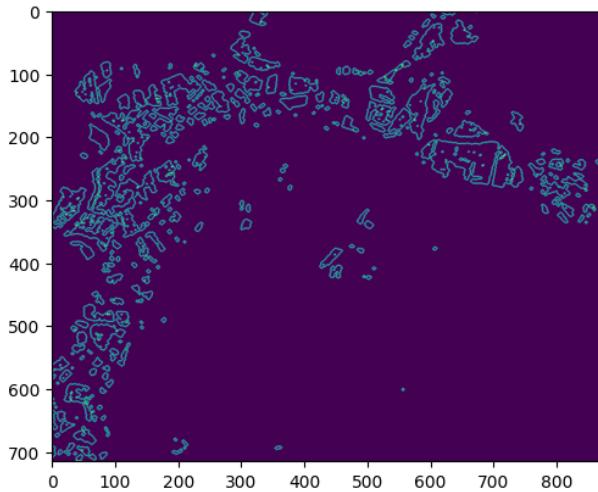


FIGURE 4 – Extraction des contours des zones inondées aux Sablons

1.7 Tâches à effectuer

Le stage a pour objectif de développer des méthodes de reconnaissance des zones inondées via des données satellites et éventuellement des drones. Les tâches spécifiques incluent :

- Valider la qualité des résultats de détection des zones inondées à partir de données Sentinel-1, en utilisant des données géolocalisées, telles que des lacs référencés dans des cartographies.
- Explorer l'utilisation de données de drones pour améliorer la détection des zones inondées. Cela impliquera d'analyser des images haute résolution pour extraire des caractéristiques spécifiques des zones touchées par les inondations.
- Explorer de nouvelles méthodes de détection, notamment les algorithmes d'apprentissage profond et les techniques d'intelligence artificielle, afin d'améliorer la précision et l'efficacité de la reconnaissance des zones inondées.

2 Méthodologie de travail

La gestion du projet est un élément crucial pour assurer la réussite de notre initiative de détection des zones inondées à l'aide de données satellitaires. Dans cette section, nous présenterons les différentes étapes et méthodologies mises en place pour atteindre nos objectifs. En définissant clairement les objectifs, les ressources nécessaires, ainsi que les risques associés, nous visons à créer un cadre solide pour guider notre équipe tout au long du processus. Une planification rigoureuse et une communication efficace entre les parties prenantes sont essentielles pour garantir le bon déroulement du projet et pour maximiser son impact dans la gestion des risques liés aux inondations. Cette approche méthodologique nous permettra d'optimiser nos efforts et de répondre efficacement aux enjeux climatiques actuels.

2.1 Organisation du Projet

2.1.1 Methodologie

Pour ce projet, nous avons adopté une méthodologie qui s'apparente aux principes de la méthode Kanban. Cette approche nous permet de gérer les flux de travail de manière efficace tout en restant flexibles et réactifs face aux besoins du projet.

La méthode Kanban est un cadre de gestion de projet visuel, souvent utilisé dans le développement de logiciels et la gestion de production. Elle se concentre sur la visualisation du flux de travail, la limitation du travail en cours (WIP), et l'amélioration continue. L'objectif principal est de garantir un flux de travail fluide et optimisé, en minimisant les goulets d'étranglement et en réduisant les temps d'attente.

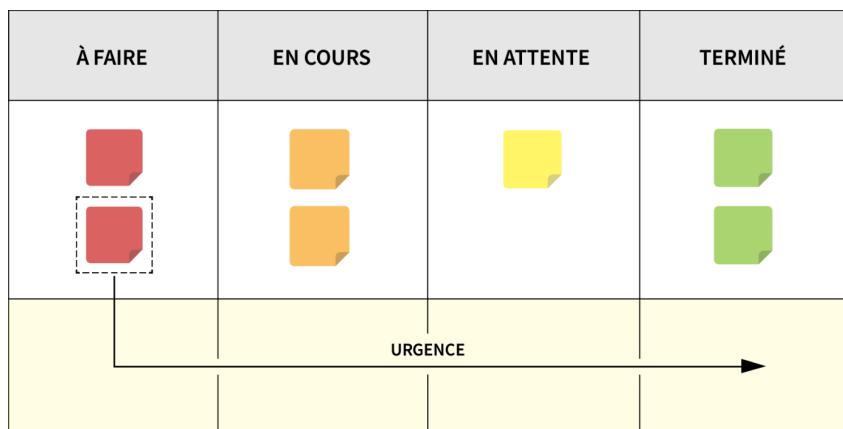


FIGURE 5 – kanban

Points Forts de la Méthode Kanban :

- Visualisation du Flux de Travail : Kanban permet de voir clairement l'état d'avancement des tâches, ce qui aide à identifier les blocages et à maintenir une transparence dans le processus de travail.
- Limitation du Travail en Cours (WIP) : En limitant le nombre de tâches en cours, Kanban aide à éviter la surcharge de travail, ce qui permet de se concentrer sur l'achèvement des tâches avant d'en commencer de nouvelles.
- Flexibilité et Adaptabilité : Kanban est flexible, permettant des ajustements en temps réel selon les priorités changeantes du projet.
- Amélioration Continue : La méthode encourage l'évaluation continue des processus, permettant des ajustements pour améliorer l'efficacité et la qualité du travail.

Notre Adaptation de la Méthode Kanban :

- Communication par E-mail : Chaque tâche est suivie par des e-mails d'avancement. Une fois une tâche terminée, un e-mail est envoyé pour résumer les progrès et indiquer si la tâche est achevée ou si des ajustements sont nécessaires.
- Objectifs Clairs et Travail Séquentiel : Nous définissons un objectif clair pour chaque tâche avant de commencer. Une fois qu'une tâche est terminée, nous passons à la suivante, évitant ainsi de gérer plusieurs tâches simultanément.
- Limitation du Travail en Cours (WIP) : Dans notre adaptation, nous veillons à ne pas sauter d'une tâche à une autre sans raison valable. Nous priorisons les tâches en fonction des besoins du projet et ne passons à une nouvelle tâche que lorsque l'objectif précédent est atteint. Cela permet de maintenir un flux de travail organisé et de minimiser les interruptions.
- Réunions de Point d'Avancement : Lorsque les principales tâches sont terminées ou lorsqu'une grande partie des tâches assignées sont accomplies, un point d'avancement est organisé pour clarifier les prochaines étapes et résoudre les éventuels problèmes.

2.1.2 Structure de l'équipe

Notre équipe est constituée de trois membres clés : un stagiaire, un superviseur et un chef d'entreprise. Chacun joue un rôle essentiel pour garantir le bon déroulement du projet :

- **Chef d'entreprise** : Responsable de la supervision générale, il veille au respect des délais et prend les décisions stratégiques pour assurer l'avancement du projet.
- **Superviseur** : Chargé de la gestion quotidienne, il définit les tâches pour le stagiaire et offre un soutien technique, ainsi que la validation des livrables.
- **Stagiaire** : Responsable de l'exécution des tâches assignées, telles que l'implémentation de code et la recherche. Le stagiaire documente chaque étape dans un rapport d'avancement, clarifiant les objectifs atteints et les responsabilités.

2.1.3 Organigramme

L'organigramme ci-dessous illustre la hiérarchie et les relations entre les membres de l'équipe.

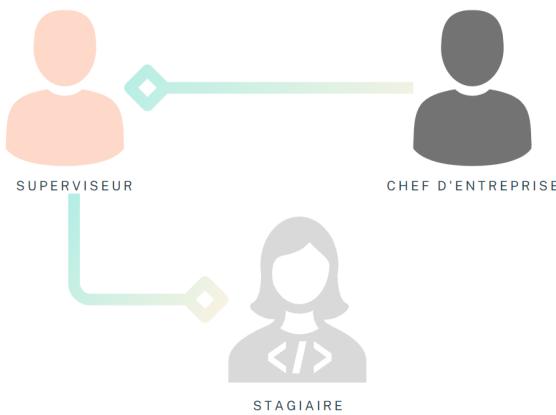


FIGURE 6 – Organigramme de l'équipe

2.2 Planning et Échéancier

2.2.1 Phases du projet

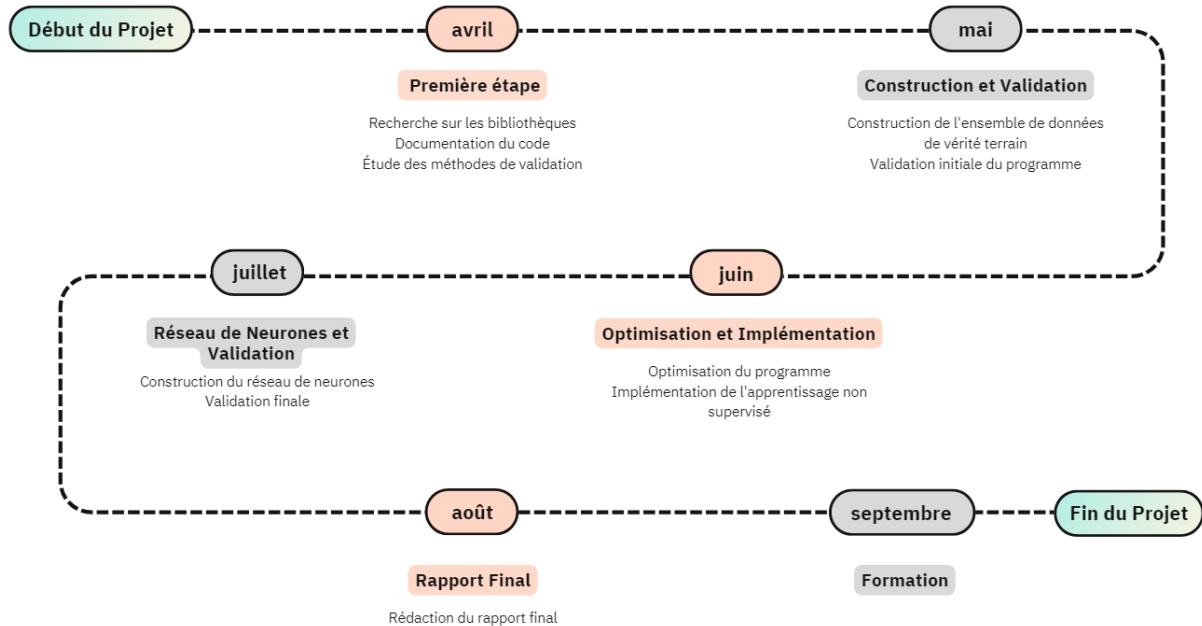


FIGURE 7 – Phases du projet

2.2.2 Diagramme de gantt

Pour détailler le déroulement du stage jusqu'à maintenant, nous avons établi un diagramme de Gantt global qui explique la période de chaque tâche

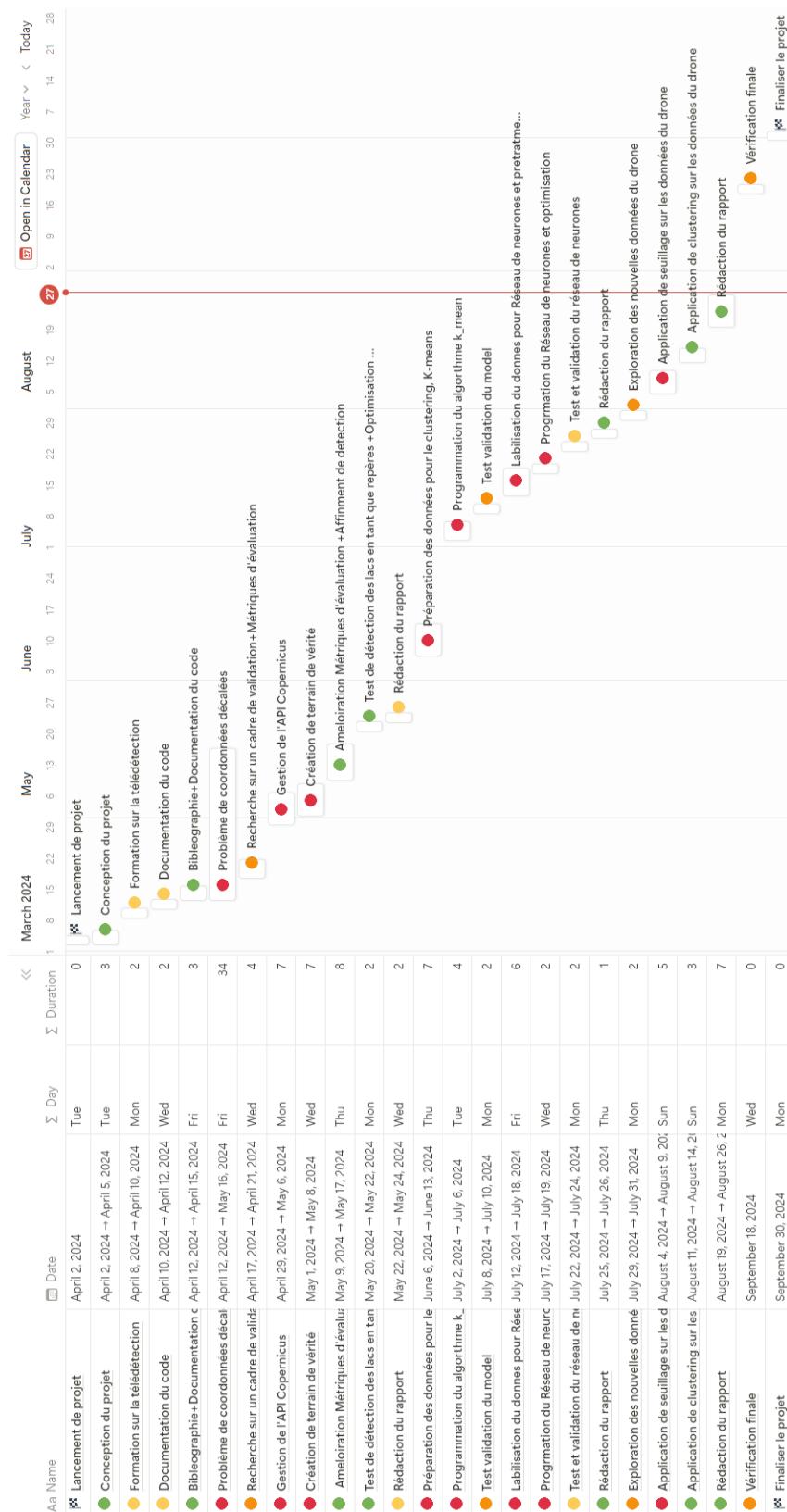


FIGURE 8 – diagramme de Gantt global

2.3 Gestion des Risques

2.3.1 Identification des risques

Dans le cadre de notre projet de détection des zones inondées à partir de données satellites, nous avons identifié plusieurs risques potentiels :

TABLE 3 – Registre des Risques

Risque	Impact	Type	Probabilité
Décalage des coordonnées	Élevé	Technique	Moyenne
Présence de coordonnées nulles	Moyen	Technique	Élevée
Manque de données de terrain	Moyen	Technique	Moyenne

- Décalage des coordonnées entre les périodes : se réfère à la différence de position géographique d'un même point ou d'une zone d'intérêt lorsqu'elle est observée à des moments différents à l'aide de données satellites. Ce phénomène peut survenir en raison de divers facteurs, tels que des erreurs dans la calibration des capteurs satellites, des variations dans l'orbite du satellite, ou des mouvements de la plateforme de capture d'image. Ce décalage peut entraîner des incohérences dans l'analyse des données, rendant difficile l'évaluation précise des changements environnementaux, tels que les inondations, entre les différentes dates d'observation.
- Présence de coordonnées nulles : Ce risque fait référence à des situations où des valeurs associées à des coordonnées géographiques spécifiques (latitude et longitude) sont manquantes ou nulles. Cela peut se produire lorsque certaines zones ne sont pas couvertes par les données, entraînant une absence d'informations pour l'analyse, ce qui complique la détection et l'évaluation des zones d'intérêt.
- Manque de données de terrain : L'absence de données de terrain pour valider nos résultats de détection constitue un défi crucial. Sans validation précise, il est difficile d'évaluer la fiabilité de notre méthode de détection. Pour juger de l'efficacité de notre programme de détection, il est essentiel d'avoir des données de référence qui représentent fidèlement les zones inondées. Cela permet non seulement d'évaluer la performance de notre algorithme, mais aussi de comparer nos résultats avec d'autres systèmes de détection. Sans un terrain de validation clair, il est impossible de distinguer les zones réellement affectées par l'eau des zones sans ambiguïté, compromettant ainsi la confiance dans nos résultats.

2.3.2 Analyse des risques

Chaque risque identifié a été analysé pour déterminer son impact potentiel sur le projet et la probabilité de son occurrence :

- Décalage des coordonnées : Impact élevé, probabilité moyenne. Ce risque pourrait entraîner des erreurs significatives dans la localisation des zones inondées.
- Présence de coordonnées nulles : Impact moyen, probabilité élevée. Les données incomplètes peuvent réduire l'efficacité de notre modèle et nécessitent des traitements supplémentaires.
- Manque de données de terrain : Impact très élevé, probabilité moyenne. L'absence de validation terrain pourrait fortement diminuer la fiabilité réelle de nos résultats.

2.3.3 Mesures de mitigation

Pour atténuer les risques identifiés, nous avons mis en place les mesures suivantes :

- Gestion du décalage des coordonnées : Malheureusement, ce problème persiste et nous n'avons pas pu le résoudre entièrement. Il reste une limitation connue de notre système, et nous continuerons à chercher des solutions potentielles dans l'avenir.
- Traitement des coordonnées nulles : Nous avons choisi d'éliminer toutes les lignes contenant des valeurs nulles. Bien que cela ait considérablement réduit la taille des données disponibles, cette approche était nécessaire, car les valeurs nulles étaient majoritaires et nous n'avions pas d'autre solution viable.
- Validation des résultats : Nous avons réussi à construire un terrain de validation à partir de l'API OSM. Cette validation a permis d'assurer la fiabilité de nos résultats de détection, augmentant ainsi la confiance dans notre méthode malgré les autres limitations.

2.4 Gestion des Ressources

2.4.1 Ressources Humaines

Dans notre projet, la gestion des ressources humaines inclut plusieurs aspects essentiels :

Recrutement et rôle

En tant que stagiaire, j'ai été recrutée pour participer au développement et à l'analyse des données. J'ai acquis des compétences spécifiques en traitement d'image et en apprentissage automatique, nécessaires pour le projet.

Formation

J'ai reçu une formation initiale sur les outils et techniques utilisés dans le projet, notamment la manipulation des données satellites et l'utilisation des logiciels de traitement d'image.

Gestion des Compétences

La supervision et le suivi de mes progrès ont été assurés par un professionnel expérimenté. Des réunions régulières ont permis de discuter des progrès réalisés, d'ajuster les tâches en fonction des besoins du projet, et d'assurer que les objectifs étaient atteints.

2.4.2 Ressources Matérielles et Financières

Pour le bon déroulement de notre projet, les ressources matérielles et financières allouées ont été les suivantes :

Ressources Matérielles

Un PC portable a été acquis spécifiquement pour ce projet, avec un coût de 1 500 euros. Cet équipement est essentiel pour le traitement des données et le développement du modèle de détection.

Le coût total du projet intègre le prix du PC portable et mon salaire mensuel pendant la durée du stage. En résumé, les ressources humaines, matérielles et financières ont été optimisées pour atteindre les objectifs du projet tout en respectant les contraintes budgétaires.

Budget Financier

Le budget du projet comprend uniquement mon salaire en tant que stagiaire, qui est de 1 400 euros par mois.

2.5 Communication et Suivi

2.5.1 Plan de Communication

Pour notre projet, qui implique une équipe réduite composée du stagiaire et de l'encadrant, nous avons adopté une approche inspirée de la méthode AGILE, bien que sans structuration en sprints, en raison de la nature de recherche et de la petite taille de l'équipe. Cette approche flexible permet de mieux répondre aux évolutions du projet et aux besoins spécifiques.

La communication au sein de notre équipe se fera principalement par les moyens suivants :

Slack	Nous utiliserons Slack pour tous les échanges quotidiens. Cette plateforme facilitera la communication rapide et la documentation des discussions, permettant de garder une trace des décisions et des points abordés.
Réunions d'Avancement	Des réunions hebdomadaires seront organisées pour discuter des progrès réalisés, des défis rencontrés, et des étapes suivantes à planifier. Ces réunions permettront également de réévaluer les priorités et d'ajuster les tâches en fonction des besoins du projet.
GitHub	Pour le partage du code et la gestion des versions, nous utiliserons GitHub. Cette plateforme assurera une collaboration harmonieuse en permettant de gérer les différentes versions du code et de faciliter le suivi des modifications apportées par chaque membre de l'équipe.

Cette stratégie de communication et de suivi vise à garantir une bonne coordination entre les membres de notre équipe, à assurer une documentation précise du travail effectué, et à faciliter la gestion du projet dans un cadre flexible et adaptable.



FIGURE 9 – Outils de communication

2.5.2 Suivi et contrôle

Dans notre projet, le planning hebdomadaire est établi par le superviseur. Chaque semaine, nous organisons une réunion pour valider les tâches en cours, suivre leur avancement et définir de nouvelles priorités. En tant que projet de recherche, notre approche est flexible ; nous n'adoptons pas un planning figé, mais nous surveillons et ajustons régulièrement notre progression pour répondre aux exigences changeantes du projet.

3 Travail réalisé

3.1 Collection de données

Notre méthode de détection nécessite des données fournies à distance, telles que des images satellites ou des données provenant de drones. Pour l'acquisition des données satellites, nous avons exploré diverses sources disponibles gratuitement et offrant une bonne précision.

3.1.1 Copernicus

Copernicus est un écosystème ouvert qui offre un accès gratuit et instantané à un grand nombre de données et de services provenant des missions Copernicus Sentinel. Il est possible d'explorer et utiliser l'imagerie satellitaire grâce au navigateur Copernicus, convivial et intuitif. On peut facilement faire des recherches,

visualiser et télécharger des données satellitaires, et bien plus encore.

Après l'établissement de l'API de connexion avec Copernicus, beaucoup de données sont disponibles que l'on peut utiliser :

Source	Caractéristiques
Sentinel-1	<ul style="list-style-type: none"> — Images radar disponibles de jour et de nuit, tous temps — Espacement des pixels : 10 m (IW), 40 m (EW) — Temps de révision : ≤ 5 jours — Disponibilité des données : Depuis octobre 2014
Sentinel-2	<ul style="list-style-type: none"> — Images à haute résolution pour surveiller la végétation, les sols, les eaux, etc. — Résolution spatiale : 10m, 20m et 60m — Délai de revisite : maximum 5 jours — Disponibilité des données : Depuis juin 2015, couverture globale depuis mars 2017
Sentinel-3	<ul style="list-style-type: none"> — Mesure de la topographie de la surface de la mer, la température, la couleur de la surface de l'océan et de la terre — Disponibilité des données : Depuis mai 2016
Sentinel-5P	<ul style="list-style-type: none"> — Mesures atmosphériques pour la qualité de l'air, l'ozone, le rayonnement UV, la surveillance et la prévision du climat — Résolution spatiale : 7 x 3,5 km — Disponibilité des données : Depuis avril 2018

3.2 Traitement d'images : seuillage + contours

Après la récupération des données de Copernicus, nous avons appliqué un traitement d'image pour détecter les zones d'eau. Ce processus se compose de deux étapes principales : le seuillage des bandes spectrales et la détection des contours.

3.2.1 Seuillage de la Bande Spectrale

Le seuillage est une technique de traitement d'image utilisée pour segmenter une image en définissant des seuils pour distinguer différentes parties de l'image. Dans notre cas, nous avons utilisé la bande VV (Vertical-Vertical) des images radar Sentinel-1 fournies par Copernicus.

Sélection de la Bande VV La bande VV est choisie en raison de sa sensibilité aux surfaces lisses et aux variations de l'eau, ce qui la rend particulièrement efficace pour détecter les zones d'eau.

Application du Seuil Un seuil initial a été fixé à 0,03 pour séparer les pixels correspondant à l'eau des autres pixels. Ce seuil est crucial pour identifier correctement les zones d'eau et sera optimisé ultérieurement pour améliorer la précision de la détection. Voici comment le seuil a été appliqué :

- **Valeurs des Pixels** : Les valeurs des pixels de l'image radar sont comparées au seuil de 0,03.

- **Classification** : Les pixels dont la valeur est inférieure ou égale à 0,03 sont classés comme zones d'eau, tandis que les pixels dont la valeur est supérieure sont classés comme non aquatiques.

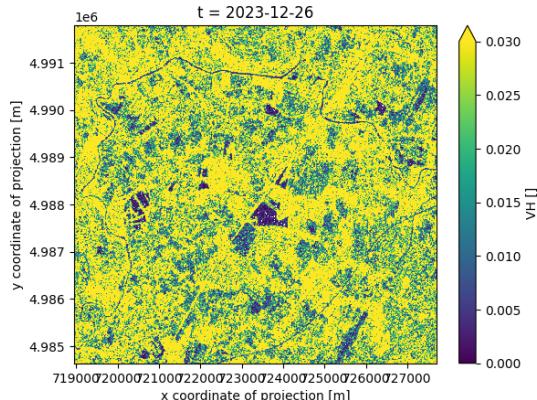


FIGURE 10 – Représentation dans la bande VH

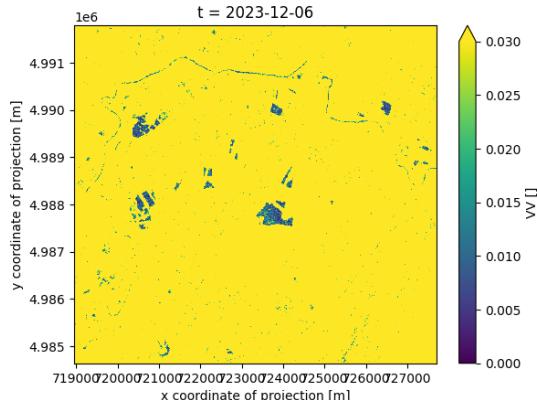


FIGURE 11 – Représentation dans la bande VV

3.2.2 Détection des Contours

Une fois le seuillage appliqué, la détection des contours est utilisée pour identifier les limites des zones d'eau. Cette étape est essentielle pour comprendre la forme et l'étendue des zones d'eau.

Inversion de l'Image L'image est inversée pour faciliter la détection des contours. Cette étape consiste à inverser les valeurs des pixels de l'image afin que les zones d'intérêt, telles que les zones d'eau, soient mises en évidence. Par exemple, si les pixels représentant l'eau ont des valeurs plus élevées, l'inversion rendra ces pixels plus sombres, ce qui simplifie la détection des contours.

Filtrage Morphologique Un filtre morphologique est appliqué pour éliminer le bruit et améliorer la qualité des contours détectés. Dans notre cas, nous utilisons la méthode de fermeture morphologique, qui consiste à appliquer successivement une opération de dilatation suivie d'une opération d'érosion. Cela permet de combler les petites lacunes dans les contours détectés et de lisser les contours pour une meilleure représentation des zones d'eau. L'élément structurant utilisé dans cette opération est en forme d'ellipse, ce qui permet d'adapter le filtrage à la forme des contours recherchés.

Détection des Contours Les contours sont détectés à l'aide de l'algorithme de détection des contours, qui trouve les contours dans une image binaire. Cet algorithme permet de tracer les frontières des zones d'eau avec précision, en identifiant les changements significatifs dans les valeurs des pixels de l'image après le seuillage et le filtrage morphologique. En appliquant cet algorithme à l'image filtrée, nous sommes en mesure de détecter les contours des zones d'eau, ce qui nous permet ensuite d'analyser la forme et l'étendue de ces zones avec précision.

3.3 Génération d'un fichier GeoJson

GeoJSON est un format de données géographiques basé sur la syntaxe JSON, utilisé pour représenter des informations géospatiales telles que des points, des lignes et des polygones avec leurs propriétés associées.

La dernière étape de la détection consiste à créer des fichiers GeoJSON en combinant les contours détectés avec les indices de nos données originales pour récupérer les coordonnées (longitude, latitude) réelles

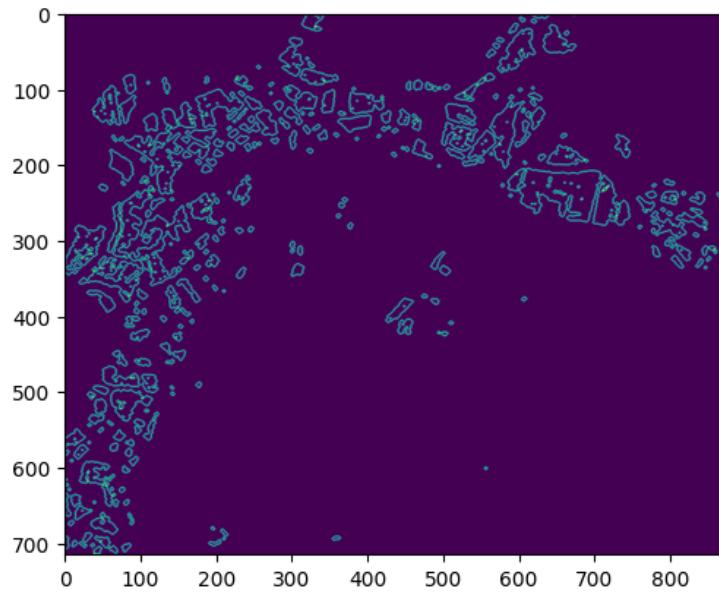


FIGURE 12 – L'extraction des contours des zones inondées aux Sablons

de chaque point dans le contour. Cette opération requiert une correspondance entre les contours extraits et les entrées dans nos données d'origine. Pour chaque contour détecté, nous parcourons les indices associés, puis récupérons les coordonnées réelles de chaque point à partir de notre jeu de données initial. Ensuite, nous construisons un fichier GeoJSON où chaque contour détecté est représenté comme une feature polygonale avec ses coordonnées géographiques. Ce fichier GeoJSON encapsule ainsi toutes les informations nécessaires pour une visualisation ou une analyse géospatiale ultérieure.

3.4 Récupération des données OSM

OSM [OpenStreetMap](#) est une carte gratuite et modifiable du monde entier. Les cartes d'OSM comprennent des données sur les routes, les adresses, les chemins de fer, les transports en commun, l'utilisation des terres et les caractéristiques naturelles, et bien plus encore, y compris la position de l'eau, y compris des cas particuliers tels que les rivières, les lacs et les océans.

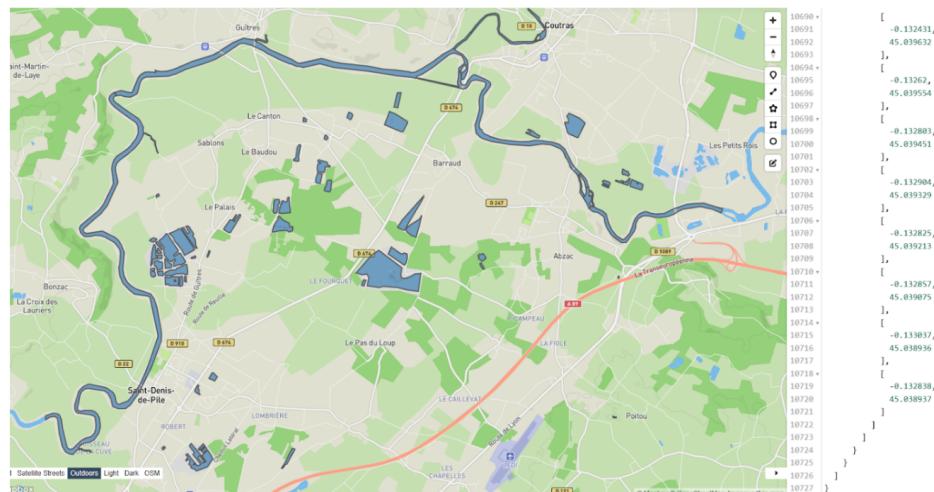


FIGURE 13 – Vrai GeoJSON représentant la zone d'eau aux Sablons

3.5 constitution d'une vérité terrain

Dans notre cas, nous devons comparer deux fichiers GeoJSON et quantifier l'approche de leurs polygones entre eux. Pour cela, nous devons nous appuyer sur des indices de calcul de similarité appropriés pour ce type de comparaison. Nous considérons les deux géométries G_1 et G_2 , où G_1 provient du premier polygon GeoJSON et G_2 du deuxième polygon. Ces indices nous permettront de mesurer la similitude entre les formes géométriques présentes dans chaque fichier, ce qui est essentiel pour comprendre les zones de recouvrement ou les différences entre les données spatiales.

Pour donner une idée, notre exemple va être la zone "Lac du Bourget" afin de visualiser la différence (Figure 14 et 15).

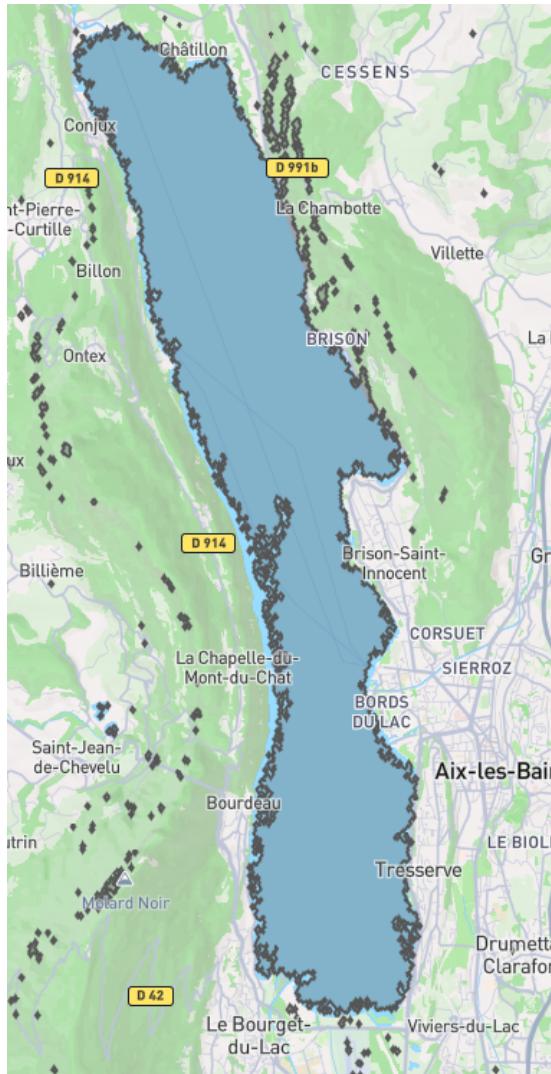


FIGURE 14 – Vrai GeoJSON dans Lac du Bourget



FIGURE 15 – GeoJSON détectée dans Lac du Bourget

3.6 Etat de l'art des mesures de correspondances de données géolocalisées

3.6.1 Indice de Jaccard

La similarité de Jaccard entre G_1 et G_2 mesure le degré de recouvrement en calculant le rapport entre l'aire de leur intersection et l'aire de leur union.[by Digitate \[2019\]](#)

Mathématiquement, cela peut être exprimé comme :

$$J(G_1, G_2) = \frac{\text{Area}(G_1 \cap G_2)}{\text{Area}(G_1 \cup G_2)}$$

3.6.2 Similarité de Dice

La similarité de Dice entre G_1 et G_2 évalue la similitude en prenant deux fois l'aire de leur intersection et en la divisant par la somme de leurs aires totales. Elle donne plus de poids aux zones de chevauchement par rapport à l'indice de Jaccard.?

Mathématiquement, cela peut être représenté comme :

$$S(G_1, G_2) = \frac{2 \times \text{Area}(G_1 \cap G_2)}{\text{Area}(G_1) + \text{Area}(G_2)}$$

3.6.3 Coefficient de Recouvrement

Le coefficient de recouvrement entre G_1 et G_2 quantifie le recouvrement en comparant l'aire de leur intersection à la plus petite de leurs aires. Il mesure à quel point les géométries se superposent les unes aux autres, en tenant compte de leur taille respective.[by Digitate \[2019\]](#)

Mathématiquement, cela peut être exprimé comme :

$$O(G_1, G_2) = \frac{\text{Area}(G_1 \cap G_2)}{\min(\text{Area}(G_1), \text{Area}(G_2))}$$

3.6.4 Indice de Sørensen-Dice

L'indice de Sørensen-Dice entre G_1 et G_2 mesure la similitude de la même manière que la similarité de Dice. Il est largement utilisé dans la comparaison de régions géographiques pour évaluer la similarité entre deux ensembles spatiaux.[Marcon \[2021\]](#)

Mathématiquement, cela peut être représenté comme :

$$S_D(G_1, G_2) = \frac{2 \times \text{Area}(G_1 \cap G_2)}{\text{Area}(G_1) + \text{Area}(G_2)}$$

3.6.5 Similarité Cosinus

La similarité cosinus entre G_1 et G_2 évalue la similitude en calculant le produit scalaire entre les deux géométries et en le divisant par le produit de leurs magnitudes. Cet indice mesure l'angle entre les vecteurs représentant les géométries. [Anonymous \[2021\]](#)

Mathématiquement, cela peut être représenté comme :

$$\text{Similarité cosinus} = \frac{G_1 \cdot G_2}{|G_1| \times |G_2|}$$

3.6.6 Distance Euclidienne

La distance euclidienne mesure la distance entre les centroïdes de G_1 et G_2 , fournissant une indication de la dissimilarité spatiale entre les géométries.

3.6.7 Distance de Hausdorff

La distance de Hausdorff évalue la dissimilarité entre G_1 et G_2 en mesurant la plus grande distance entre un point de l'une des géométries et le point le plus proche de l'autre géométrie. Elle capture les variations extrêmes entre les deux ensembles spatiaux.[Techno-Science \[2021\]](#)

3.7 Validation de la détection

L'utilisation de chaque indice va donner une matrice d'intra-classe qui calcule la similarité entre chaque polygone du GeoJSON 1 avec les polygones du GeoJSON 2

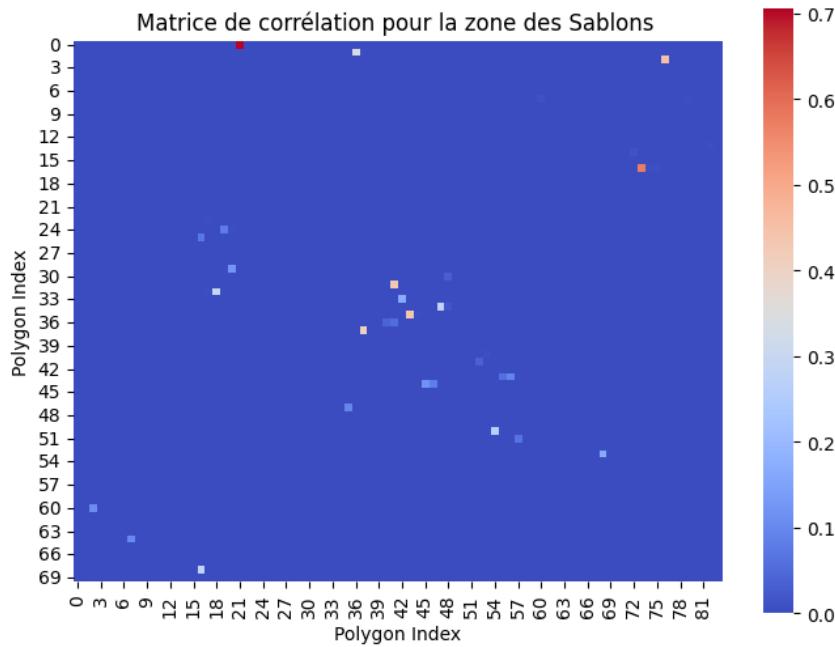


FIGURE 16 – Exemple de matrice intra-class

3.8 Amélioration des algorithmes de détection

3.8.1 augmentation de la précision

Un des problèmes rencontrés dans la détection est lorsque des polygones se trouvent à l'intérieur d'autres polygones, ce qui peut affecter le processus de validation.

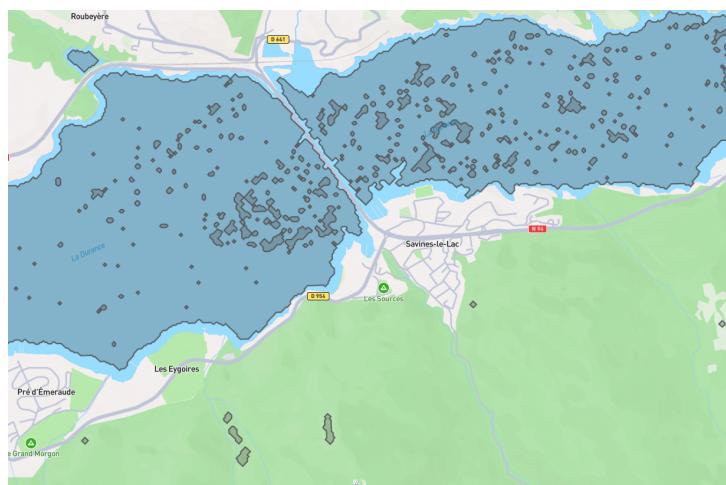


FIGURE 17 – problem de detection

Pour résoudre le problème de détection qui va affecter la quantification de précision de notre détection, on ajoute une autre phase : une fusion basée sur la morphologie d'ensemble. Elle consiste à fusionner

tous les polygones et, par conséquent, cela nous guidera à éliminer les polygones inclus dans d'autres polygones

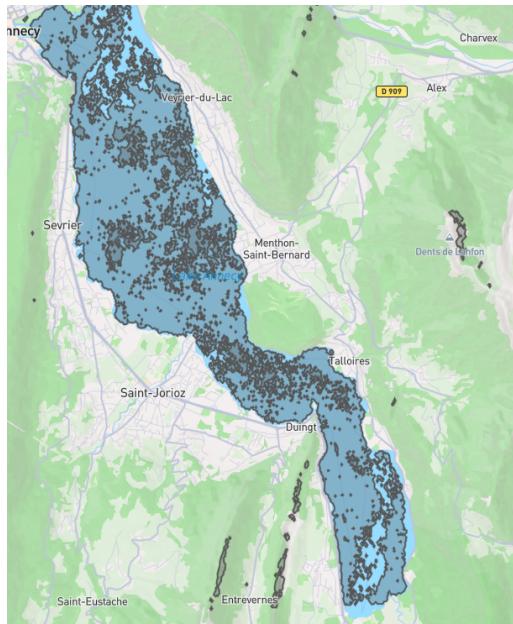


FIGURE 18 – Représentation dans la bande VH

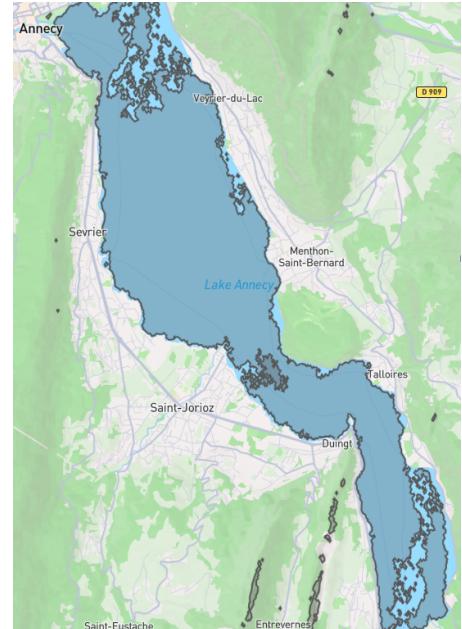


FIGURE 19 – Représentation dans la bande VV

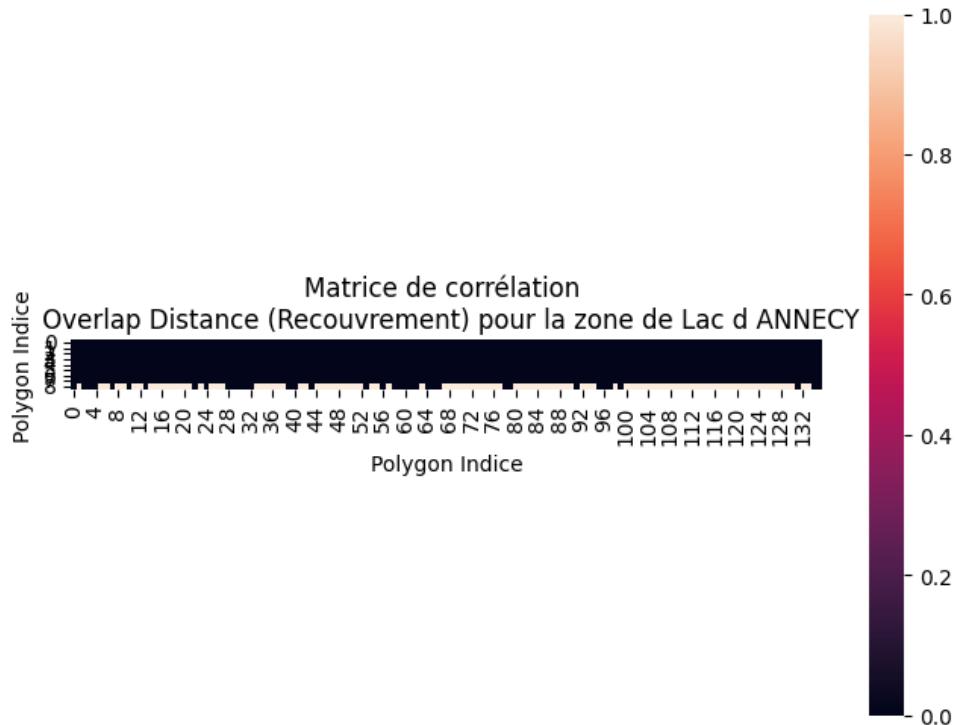


FIGURE 20 – Matrice de intra-class de lac Annecy

3.8.2 L'optimisation des hyperparamètres

Notre objectif est d'optimiser les hyperparamètres tels que le seuil de bande, le coefficient de dilatation et le nombre d'itérations de dilatation. Pour simplifier le processus de jugement, nous avons décidé

de sélectionner le plus grand polygone existant dans la zone définie des deux côtés, une approche particulièrement pertinente étant donné que nous travaillons sur les lacs.

```
Indice de Jaccard entre les deux plus grands polygones: 0.7869777167946205
Indice de dice entre les deux plus grands polygones: 0.880791863713059
Indice de overlap entre les deux plus grands polygones: 0.9998327080126596
Indice de sorensen entre les deux plus grands polygones: 0.880791863713059
Indice de euclidean entre les deux plus grands polygones: 0.004569952512308837
Indice de hausdorff entre les deux plus grands polygones: 0.0
```

FIGURE 21 – Résultat obtenu des distances entre le plus grand polygone (réalité vs détection) du lac d'Annecy.

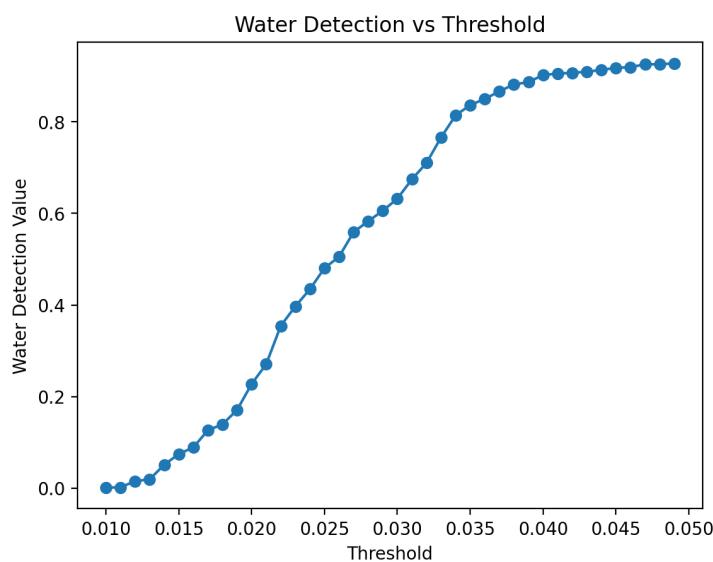


FIGURE 22 – Variation du seuil par rapport à la qualité de détection

Afin d'améliorer notre programme de détection, nous avons entrepris une optimisation classique des hyperparamètres. L'objectif est de trouver les meilleures hyperparamètres qui contrôlent la détection en explorant les différentes valeurs possibles.

Nous avons commencé par ajuster le seuil de détection, en testant diverses valeurs jusqu'à obtenir le meilleur score possible. Parallèlement, nous avons optimisé l'épaisseur de l'opération de fermeture, un processus incluant la dilatation et l'érosion dans le traitement d'image. Cette opération a été ajustée de manière similaire, jusqu'à ce que nous trouvions les meilleures hyperparamètres pour notre programme à ce niveau.

3.9 Amélioration de la Détection des Zones d'Eau par Clustering (Algorithm K-Means)

Pour explorer de nouvelles options et améliorer notre programme, nous avons décidé de remplacer la phase de suillage des bandes par une méthode plus innovante : le clustering. Plus précisément, nous avons opté pour l'algorithme K-Means classique pour accomplir cette tâche.

K-Means est un algorithme de partitionnement qui divise les données en k clusters distincts, où chaque point de données appartient au cluster avec la moyenne la plus proche. Dans notre projet, nous avons appliqué K-Means sur les bandes VV et VH normalisées des images radar pour améliorer la détection des zones d'eau.

Nous avons d'abord supprimé les valeurs nulles pour éviter les biais dans le clustering. Ensuite, nous avons utilisé les données nettoyées pour ajuster le modèle K-Means avec un nombre de clusters défini à 2, correspondant aux catégories "eau" et "non eau". Les labels résultants ont été reformatés pour correspondre à la forme originale des images .



FIGURE 23 – Représentation réelle du lac Bourget

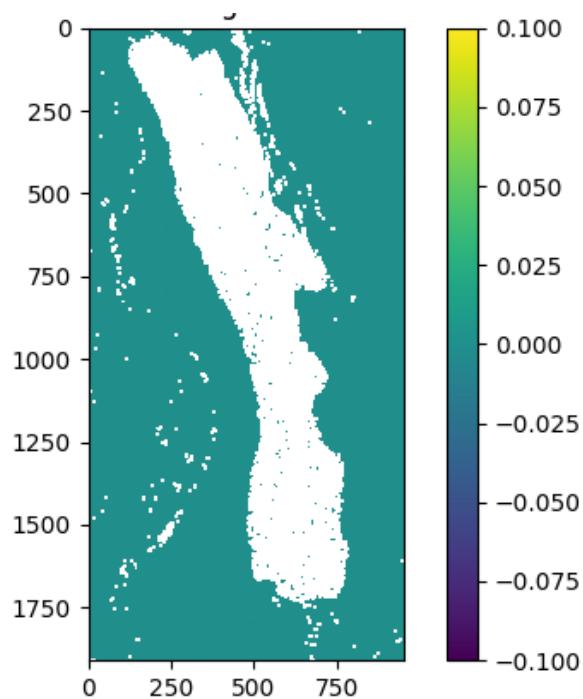


FIGURE 24 – Représentation du lac (par algorithme k-means) basée sur les bandes VH et VV

D'après les résultats obtenus grâce au clustering, nous avons intégré cette méthode dans notre programme pour construire des résultats sous forme de polygones. À la fin du processus, après le regroupement des polygones, nous avons obtenu un résultat satisfaisant. Notre programme a réussi à détecter les zones d'eau dans le lac efficacement, et nous avons pu distinguer une différence notable entre la qualité produite par le seuillage classique et notre méthode de clustering.

Bien que le seuillage classique soit une idée simple, il reste efficace, simple et rapide. Même en validation qualitative, le seuillage classique conserve un score élevé par rapport au clustering. Cependant, le clustering apporte une approche innovante qui peut être améliorée pour surpasser les méthodes traditionnelles.



FIGURE 25 – Résultat final de l’algorithme k-means comparé avec le vrai lac Bourget

3.10 Amélioration de la Détection des Zones d’Eau par Réseau de Neurones

Dans le cadre de notre projet, une approche innovante que nous pouvons intégrer est l’utilisation de réseaux de neurones. Nous avons divisé ce processus en trois phases principales : la préparation des données d’entraînement, l’entraînement du modèle, et l’évaluation du modèle.

3.10.1 Préparation des Données d’Entraînement

La première étape consiste à préparer les données nécessaires pour entraîner le réseau de neurones. Nous utilisons l’API d’OpenStreetMap (OSM) pour collecter des données pertinentes en configurant l’API avec les coordonnées géographiques des zones ciblées.

Pour chaque pixel, les bandes VH (Vertical-Horizontal) et VV (Vertical-Vertical) sont concaténées pour fournir des valeurs d’entrée. Les données OSM permettent de distinguer les pixels représentant de l’eau (valeur 1) de ceux n’en contenant pas (valeur 0). Ensuite, les données sont nettoyées pour éliminer les valeurs manquantes (NaN). Les images sont ensuite normalisées et redimensionnées afin qu’elles puissent être traitées efficacement par le réseau de neurones. Ces étapes sont cruciales pour garantir la qualité et la fiabilité du jeu de données d’entraînement, ce qui est essentiel pour entraîner notre modèle de détection de manière optimale.

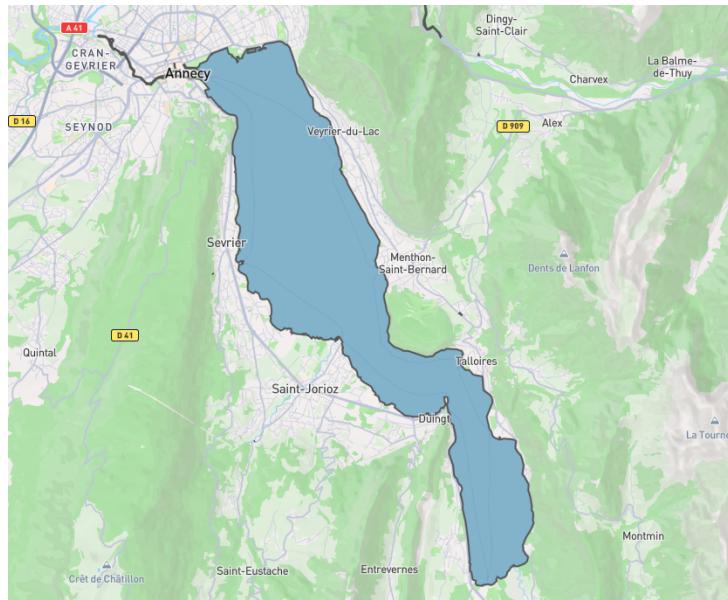


FIGURE 26 – Données effectives du lac d'Annecy

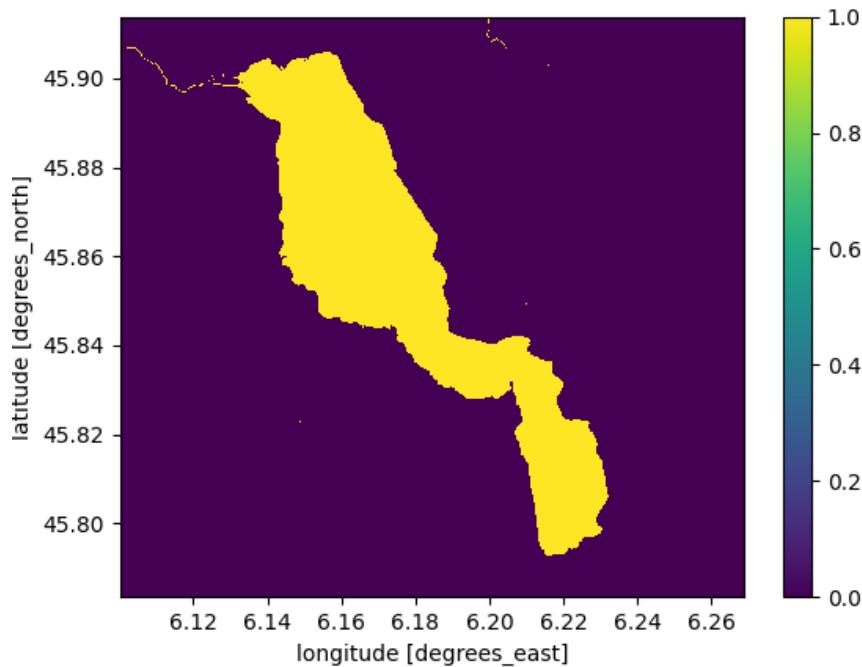


FIGURE 27 – Données labellisées du lac d'Annecy

3.10.2 Entraînement du Modèle

Une fois les données prêtes, nous passons à la phase d’entraînement du modèle. Nous utilisons les données préparées pour entraîner un réseau de neurones, en ajustant les poids et les biais du modèle pour minimiser l’erreur de prédiction. Cela implique l’utilisation d’algorithmes d’apprentissage supervisé, tels que la rétropropagation, et peut nécessiter plusieurs itérations pour obtenir des résultats optimaux.

Pour maintenir la simplicité, étant donné que notre problème n’est pas extrêmement complexe, nous avons créé un modèle de réseau de neurones simple avec deux couches principales. La première couche est une

couche dense avec 16 neurones et une fonction d’activation ReLU (Rectified Linear Unit), prenant en entrée deux valeurs correspondant aux bandes VH et VV. La deuxième couche est une couche dense avec un seul neurone et une fonction d’activation sigmoïde, utilisée pour produire une sortie binaire indiquant si le pixel représente de l’eau ou non.

Nous utilisons 75 pour cent des données totales pour l’entraînement du modèle et réservons les 25 pour cent restants pour tester et évaluer la performance du modèle. Cette approche nous permet de vérifier l’efficacité du modèle sur des données non vues pendant l’entraînement, garantissant ainsi une évaluation précise de ses performances.

3.10.3 Évaluation du Modèle

Ensuite, nous avons évalué les performances du modèle en utilisant les 25 pour cent des données réservées à cet effet. Les résultats de l’évaluation comprennent la perte et l’exactitude du modèle sur les données de test. Ces métriques sont essentielles pour comprendre l’efficacité et la précision de notre modèle.

Pour approfondir l’évaluation, nous avons tracé la courbe ROC (Receiver Operating Characteristic). Cette courbe illustre la capacité du modèle à distinguer entre les classes en montrant la relation entre le taux de faux positifs (FPR) et le taux de vrais positifs (TPR). Une courbe ROC proche de la diagonale représente une prédiction aléatoire, tandis qu’une courbe plus proche du coin supérieur gauche indique une meilleure performance du modèle.

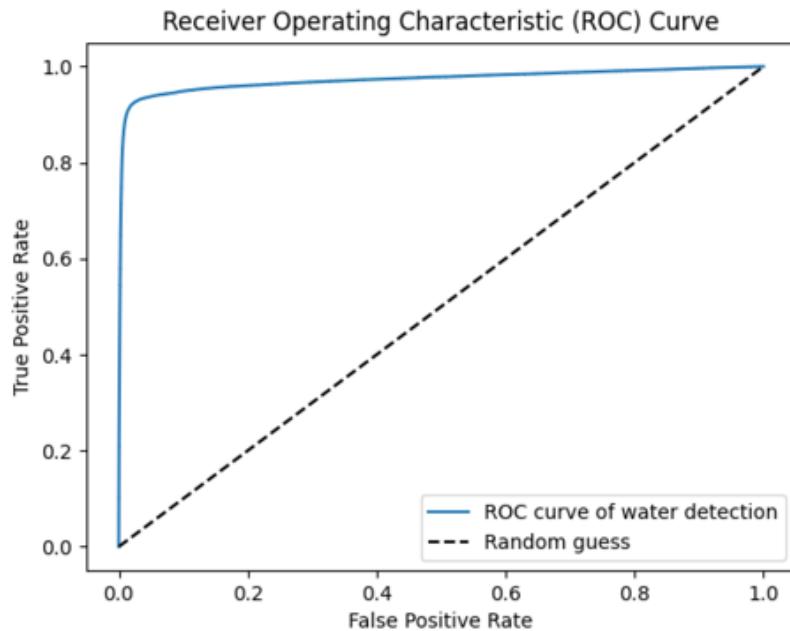


FIGURE 28 – Évaluation ROC du modèle

La courbe ROC obtenue démontre que notre réseau de neurones a bien adressé le problème de détection des zones d’eau. En effet, la courbe couvre une grande partie de la surface sous la courbe ROC, ce qui est un indicateur de bonne performance. De plus, une évaluation simple de l’exactitude du modèle a révélé une précision de 97 %.

En résumé, la courbe ROC et la haute précision obtenue indiquent que notre modèle de réseau de neurones est performant et adapté pour la détection des zones d’eau dans les images analysées.

3.11 Exploration de données du Drone

Les nouvelles données que nous avons reçues proviennent de drones et sont constituées d'images JPEG avec une grande résolution et des détails clairs, ce qui permet une exploitation efficace. Ces images sont presque parfaites et ne sont pas affectées par les changements climatiques. Cependant, la quantité de ces images est limitée à quatre, accompagnées d'une représentation 3D de la zone. Malheureusement, les images ne sont pas labellisées et il n'y a aucun système de coordonnées (longitude, latitude) pour identifier précisément l'emplacement géologique de cette zone, ce qui représente un défi pour leur exploitation.



FIGURE 29 – Main caption for all figures

3.11.1 Préparation et Traitement des Images

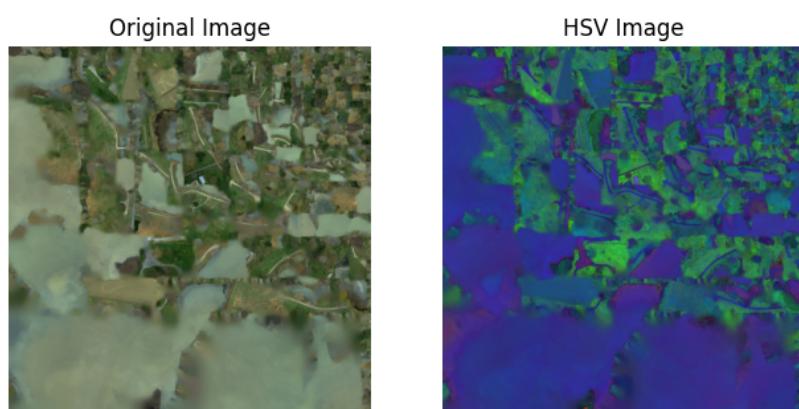


FIGURE 30 – transformation HSV

La première étape que nous avons prise pour traiter ces images consiste à nous familiariser avec notre

algorithme précédent, qui inclut le seuillage, le réseau de neurones et l'algorithme K-means. Pour préparer cette étape, nous avons décidé de changer la représentation colorimétrique des images JPEG en HSV (Hue, Saturation, Value). Cette représentation est plus adaptée à la détection, car elle sépare la luminance (intensité lumineuse) des informations de couleur. En utilisant l'espace colorimétrique HSV, nous pouvons mieux identifier et extraire les caractéristiques pertinentes des images, ce qui améliore notre capacité à détecter des objets ou des anomalies dans les données.

3.11.2 Seuillage sur les Canaux H et S

Une fois que nous avons converti les images en espace colorimétrique HSV, nous avons appliqué le seuillage sur le canal de teinte (H) et le canal de saturation (S) afin de distinguer entre deux classes : "eau" et "pas d'eau". En définissant des seuils appropriés pour ces canaux, nous avons pu créer une image segmentée où les zones représentant de l'eau sont mises en évidence. Ce processus de seuillage permet de faciliter l'analyse ultérieure des zones inondées, en mettant en lumière les régions d'intérêt tout en minimisant les influences des variations de lumière et des textures non pertinentes dans l'image.

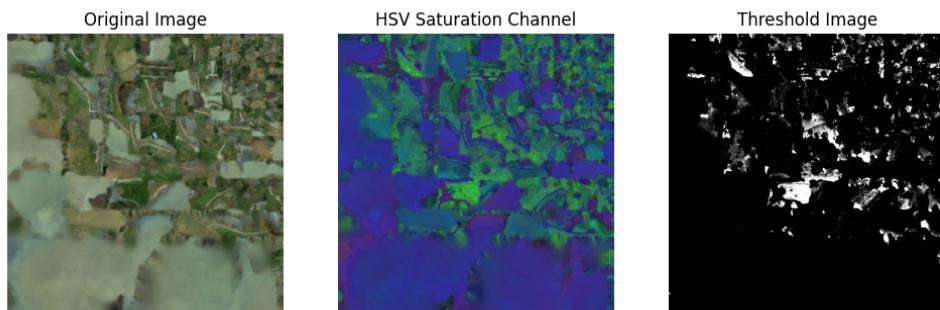


FIGURE 31 – Application du seuillage classique sur les données Drone

3.11.3 Application de l'Algorithme K-Means

Après avoir préparé les données en utilisant la représentation HSV, nous avons appliqué un algorithme non supervisé, K-Means, pour segmenter l'image en deux clusters. Cette méthode permet de regrouper les pixels en différentes classes en fonction de leur couleur et de leur intensité. En tirant parti des caractéristiques visuelles de l'image dans l'espace HSV, nous avons pu identifier distinctement les zones d'eau et les autres éléments présents. L'algorithme K-Means a ainsi divisé l'image, facilitant la détection des régions d'intérêt, même sans données étiquetées. Cette approche a permis de mieux distinguer les zones d'eau des autres surfaces, augmentant ainsi la précision de notre analyse. En résumé, la combinaison de la représentation HSV et de l'algorithme K-Means a été cruciale pour améliorer notre capacité à segmenter les images et à identifier les zones pertinentes pour l'étude des inondations.

Malheureusement, en raison de l'absence de données étiquetées pour nos images, nous ne pouvons pas appliquer un apprentissage supervisé avec un réseau de neurones. Les réseaux de neurones nécessitent des données d'entraînement étiquetées pour apprendre à distinguer efficacement les différentes classes d'objets dans une image. Cette limitation nous empêche d'exploiter pleinement le potentiel de ces

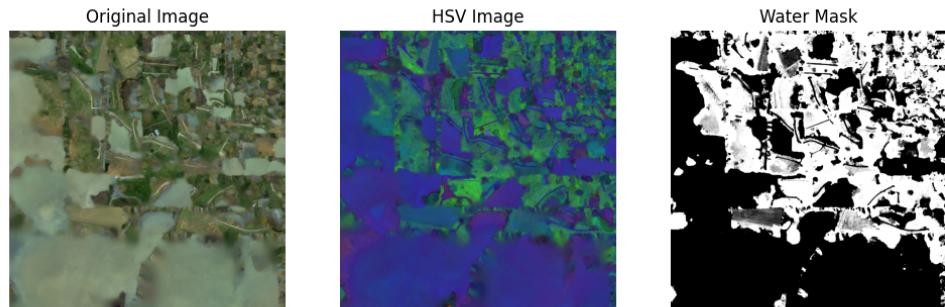


FIGURE 32 – Application de K-Means sur les données Drone

modèles avancés dans notre analyse des données. Cependant, une fois que des données étiquetées seront disponibles, l'application d'un réseau de neurones pourrait considérablement améliorer la précision de la détection des zones d'intérêt dans les images.

3.12 Limites de l'Étude

3.12.1 Problème du décalage

Comme on peut le voir dans la figure, une comparaison des coordonnées pour la même zone avec seulement une semaine de différence révèle un décalage significatif. Il est clairement visible que la position du cours d'eau a changé, ce qui est impossible

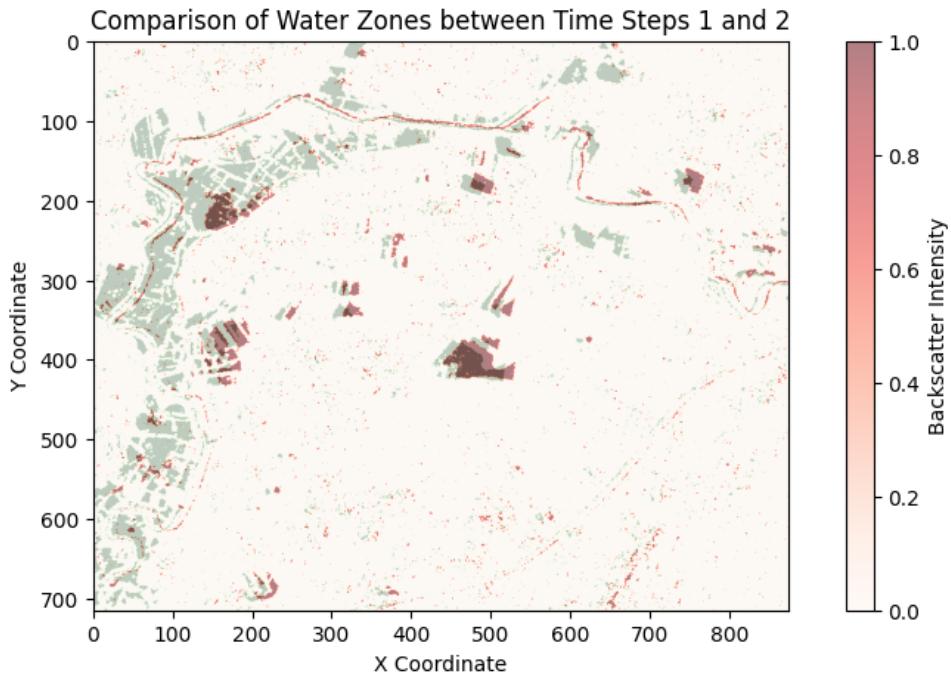


FIGURE 33 – Problème de décalage de coordonnées

Nous avons essayé de changer tout le code qui fait la demande de l'API et de le remplacer par une demande de l'API avec la bibliothèque SentinelHub, comme le montre la figure. Cependant, à la fin, nous avons trouvé des données, mais nous n'avons pas réussi à récupérer les coordonnées (longitude, latitude) correspondant à nos données comme avec l'ancien code.

Nous avons remarqué que ce problème se manifeste uniquement dans la zone du Sablon en France, lorsqu'on compare les requêtes obtenues avec beaucoup d'autres zones. Pour le moment, nous pouvons considérer ces données comme erronées et les mettre de côté afin de ne pas perdre plus de temps.

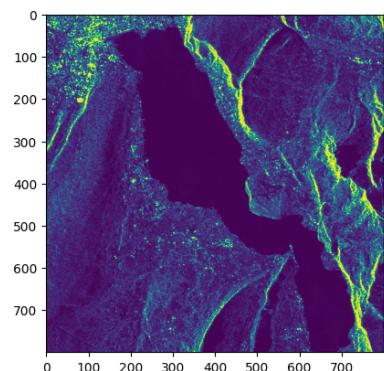


FIGURE 34 – sentinelhub Données

Conclusion Générale

En conclusion, ce projet a constitué une expérience exceptionnelle et formatrice qui m'a permis de me plonger au cœur des enjeux liés à la détection des inondations, un problème environnemental majeur de notre époque. Grâce à l'utilisation avancée des données satellitaires, nous avons pu développer des solutions innovantes pour améliorer la précision et la rapidité de la détection des inondations, essentielles pour protéger les vies et les biens.

Au cours de cette période, j'ai acquis une multitude de compétences techniques en traitement d'images et en apprentissage automatique. J'ai également eu l'opportunité de me familiariser avec des outils et des méthodologies modernes, tels que la méthode Kanban, qui a grandement contribué à l'organisation et à l'efficacité de notre travail d'équipe. La structure flexible que cette méthode offre a permis une gestion dynamique des tâches, facilitant ainsi une réévaluation constante des priorités en fonction des besoins du projet.

Un aspect particulièrement enrichissant de ce projet a été l'interaction et la collaboration avec des professionnels expérimentés. Les échanges réguliers avec mon encadrant, Monsieur Arnaud Revel, et mes collègues ont été essentiels pour surmonter les défis rencontrés tout au long de ce processus. Leurs retours constructifs m'ont permis d'améliorer mes compétences et d'approfondir ma compréhension des enjeux techniques et stratégiques du projet.

Cependant, ce projet n'a pas été sans défis. La variabilité des données satellitaires en fonction des conditions climatiques et la complexité de l'intégration des nouveaux systèmes dans les infrastructures existantes ont nécessité une réflexion approfondie et des ajustements constants. Ces obstacles ont non seulement testé notre capacité d'adaptation, mais ont également mis en lumière l'importance d'une approche méthodologique rigoureuse et d'une communication ouverte au sein de l'équipe.

À l'avenir, les solutions développées dans le cadre de ce projet pourraient avoir un impact significatif sur la gestion des catastrophes naturelles. En adaptant ces technologies aux spécificités régionales et en établissant des partenariats avec des compagnies d'assurance, nous pouvons espérer élargir leur portée et leur efficacité.

En somme, ce projet a non seulement contribué à mon développement professionnel et personnel, mais il a également renforcé ma conviction que l'innovation technologique est cruciale pour relever les défis environnementaux contemporains. Je suis motivé à continuer à explorer ces domaines et à apporter ma contribution à des initiatives visant à améliorer notre résilience face aux catastrophes naturelles. Je suis reconnaissant pour l'opportunité qui m'a été offerte et j'attends avec impatience de futures collaborations qui continueront à avoir un impact positif sur notre société.

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude à Monsieur Arnaud Revel, mon encadrant de stage, pour son accompagnement exceptionnel tout au long de cette aventure. Sa patience, sa stratégie pédagogique et son expertise ont été des catalyseurs essentiels à mon apprentissage et à ma réussite.

Je souhaite également remercier chaleureusement Monsieur David Helbert pour ses précieuses remarques académiques et ses conseils avisés, qui ont grandement contribué à enrichir ma compréhension et mon développement professionnel.

Un remerciement spécial est adressé à Monsieur Clency Perrine, dont le soutien mental et émotionnel a été un pilier dans les moments les plus exigeants de mon stage.

Enfin, je suis profondément reconnaissant envers notre chef d'entreprise, Monsieur Jerome Petit, pour son soutien indéfectible et sa constante bienveillance. Son engagement envers le bien-être de ses employés et son leadership inspirant ont grandement marqué mon expérience au sein de l'entreprise.

Ces remerciements s'adressent aussi à tous mes collègues et à toutes les personnes qui ont participé, de près ou de loin, à cette expérience enrichissante. Votre aide et votre gentillesse ont été des atouts précieux qui ont pavé mon chemin vers la réussite de ce stage.

Références

- Anonymous. Cosine similarity : How does it measure similarity ?, 2021. URL <https://ichi.pro/fr/similitude-cosinus-comment-mesure-t-il-la-similitude-les-mathematiques-derriere-et-l-utilisation>. Accessed : 2024-08-28.
- Ignio by Digitate. Similarity coefficients : A beginner's guide to measuring string similarity, 2019. URL <https://medium.com/@igniobydigitate/similarity-coefficients-a-beginners-guide-to-measuring-string-similarity-d84da77e8c5a>. Accessed : 2024-08-28.
- Ministère de la Transition écologique et de la Cohésion des territoires. Savoir sur les pluies intenses et les inondations en france, 2023. URL <https://www.ecologie.gouv.fr/dossiers/comprendre-risques-naturels-sen-proteger/savoir-pluies-intenses-inondations-france>.
- Centre de ressources pour l'adaptation au changement climatique. Impacts des inondations et mesures d'adaptation, 2023. URL <https://www.adaptation-changement-climatique.gouv.fr/dossiers-thematiques/impacts/inondation>.
- Office français de la biodiversité (OFB). Questions-réponses sur les inondations, 2023. URL <https://www.ofb.gouv.fr/questions-reponses-sur-les-inondations>.
- Jiaxin Li, Ronghua Ma, Zhigang Cao, Kun Xue, Junfeng Xiong, Minqi Hu, and Xuejiao Feng. Satellite detection of surface water extent : A review of methodology. *Water*, 14(7) :1148, 2022. doi:[10.3390/w14071148](https://doi.org/10.3390/w14071148).
- Eric Marcon. Beta diversity indices in pairs, 2021. URL <https://ericmarcon.github.io/MesuresBioDiv2/chap-BetaPaires.html>. Accessed : 2024-08-28.
- Techno-Science. Distance de hausdorff, 2021. URL <https://www.techno-science.net/definition/5199.html>. Accessed : 2024-08-28.